arXiv:0812.0191v1 [q-bio.SC] 1 Dec 2008

Calculation of the relative metastabilities of proteins in subcellular compartments of *Saccharomyces cerevisiae*

Jeffrey M. Dick

Department of Earth and Planetary Science
University of California
Berkeley, CA 94720

Background: The distribution of chemical species in an open system at metastable equilibrium can be expressed as a function of environmental variables which can include temperature, oxidation-reduction potential and others. Calculations of metastable equilibrium for various model systems were used to characterize chemical transformations among proteins and groups of proteins found in different compartments of yeast cells.

Results: With increasing oxygen fugacity, the relative metastability fields of model proteins (including isoforms of glutaredoxin and thioredoxin, and compartmental proteomes) for major subcellular compartments go as mitochondrion, endoplasmic reticulum, cytoplasm, nucleus. Compared with experimental determination of redox potential (Eh) in these compartments, the order of the endoplasmic reticulum and nucleus is swapped. In a metastable equilibrium setting at relatively high oxygen fugacity, proteins making up actin are predominant, but those constituting the microtubule occur with a low chemical activity. Nevertheless, interactions of the microtubule with other subcellular compartments are essential in cell development. A reaction sequence involving the microtubule and spindle pole proteins was predicted by combining the known intercompartmental interactions with a hypothetical program of oxygen fugacity changes in the local environment. In further calculations, the most-abundant proteins within compartments generally occur in relative abundances that only weakly correspond to a metastable equilibrium distribution. However, physiological populations of proteins that form complexes often show an overall positive or negative correlation with the relative abundances of proteins in metastable assemblages.

Conclusions: This study explored the outlines of a thermodynamic description of chemical transformations among interacting proteins in yeast cells. Full correspondence of the model with biochemical and proteomic observations was not apparent, but the results suggest that these methods can be used to measure the degree of departure of a natural biochemical process or population from a local minimum in Gibbs energy.

Author Summary

Part of a cell's expenditure of metabolic fuel is directed toward the formation of proteins, including their synthesis and transport to other compartments. Even when it is normalized to the lengths of the proteins, the energy required for protein formation is not a constant, but depends on the composition and environment of the protein. If these energy differences are quantified, the relative abundances of model proteins in metastable equilibrium can be calculated. The compositions of these metastable assemblages depend on local environmental variables such as oxygen fugacity, which is a scale for oxidation-reduction potential in a system. I calculate the oxygen fugacities for equal chemical activities of model proteins in intercompartmental interactions and use the results to obtain model values of oxygen fugacity for subcellular compartments. I show that a environmental gradient of oxygen fugacity can potentially drive the formation of proteins in a sequential order determined by their chemical compositions and Gibbs energies. I also show that the relative abundances of proteins within compartments and of those that form complexes have a dynamic range that can be approximated in some metastable equilibrium assemblages. These results provide theoretical constraints on the natural emergence of spatial and temporal patterns in the distributions of proteins and imply that work done by maintaining oxidation-reduction gradients can selectively alter the degree of formation of proteins and complexes.

Introduction

Subcellular compartmentation is a basic feature of eukaryotic life [1, 2, 3, 4]. There exist in eukaryotic cells gradients between subcellular compartments of chemical properties such as pH [5, 6, 7, 8], oxidation-reduction or redox state [9, 10, 11, 12] and chemical activity of water [13, 14, 15, 16]. Furthermore, the proteins required by yeast and other organisms are unevenly localized throughout cells [17, 18, 19, 20, 21, 22, 23]. Even within compartments or among the proteins that interact to form complexes, the relative abundances or levels of different proteins are not equal [24, 25, 26], and different proteins predominate in the various subcellular populations depending on growth state of the cell [27, 28], and exposure to environmental stress [29, 30, 31, 32].

Much attention has been given to the use of thermodynamics in describing and understanding driving forces in biological evolution. Energy minimization imparts a direction for spontaneous change of a system, and response of a system in this direction can at times be tied to an increase in relative fitness [33, 34, 35, 36, 37]. A biological system that moves away from minimum energy does not break the laws of thermodynamics but couples its endergonic reactions with the exchange of matter and energy in its surroundings [38, 39, 40, 41]. The thermodynamic characteristics of open systems are thus of particular interest to biological evolution [42, 43, 44]; in particular, the interactions of organisms with their environments are important influences on the stable compositions and distributions of genes or organisms [45, 46, 47, 48].

Why are proteins not equally distributed inside cells? Physical separation of key enzymes is thought to be essential in the cytoskeletal network and in regulation of metabolic pathways and other cellular functions [49, 50, 51]. The patterns of subcellular structure persist even though populations of proteins turnover through continual degradation and synthesis in cells [52, 53, 54, 55], and despite the endergonic, or energy-consuming, qualities of protein biogenesis [56, 57]. It can be shown that the relative abundances of amino acids in proteins correlate inversely with the metabolic cost of synthesis of the amino acid [58, 59], which is a temperature-dependent function [60]. The starting premise of this study, then, is that protein formation reactions are unfavorable to different degrees, depending on the environments and compositions of the biomolecules.

The application of equilibrium chemical thermodynamics as a way to characterize the relative stabilities of minerals as a function of temperature, pressure and oxidation-reduction potential [61, 62, 63], or to calculate the relative abundances of coexisting inorganic [64, 65] and/or organic species [41, 66], is well documented in the geochemical literature. An advantage of performing quantitative chemical thermodynamic calculations for many different model systems is that the equilibrium state serves as a frame of reference for describing both reversible and irreversible chemical changes. For example, the weathering of igneous rocks is an overall irreversible process but the sequences of minerals formed can nevertheless be predicted after initial formulation of the relative stability limits of the chemical species involved [67, 68]. One of the motivations for this study is to see whether a similar approach could be used to describe the sequence of events in irreversible subcellular processes.

The thermodynamic calculations reported in this study are based on algorithms for calculating the standard molal Gibbs energies of ionized proteins [69] and a chemical reaction framework that is used to compute metastable equilibrium relative abundances of proteins [70]. The Supporting Information for this paper includes the software package (Text S1) and the program script and data files (Text S2) used to carry out these calculations. The theoretical approach adopted here is based on the description of a chemical system in terms of intensive variables. These variables are temperature, pressure and the chemical potentials of the system. It is convenient to denote the chemical potentials by the chemical activities or fugacities of basis species, for example the activity of H^+ (which defines pH) or the fugacity of oxygen. This permits comparison of the parameters of the model with reference systems described in experimental and other theoretical biochemical studies.

A few notes on terminology follow. *Formation* of a protein refers to the overall process of protein biosynthesis and translocation to a specific compartment. *Activity* and *species* denote, respectively, chemical activity and chemical species, not enzyme activity or biological species. In the present study, activity coefficients are taken to be unity, so the chemical activities are equivalent to molal concentrations. Below, *oxidation-reduction potential* and oxygen fugacity are used synonymously, and *redox* refers specifically to Eh. The oxidation-reduction potential of a system can be expressed in terms of Eh using an equation given in the Methods. The overall compositions of proteins in compartments are referred to here as *proteologs* (or model proteologs). The *interactions* of proteins are processes in which the proteins come into physical contact, for example in transport processes between compartments and in the formation of complexes. If a process results in a change in the composition of a population of interacting proteins, then a chemical *reaction* has occurred. Protein-protein interactions do not necessarily correspond to chemical reactions. However, a population of interacting proteins does chemically react

Table 1: Subcellular isoforms of glutaredoxin, thioredoxin and thioredoxin reductase in yeast[a].

Protein	SWISS-PROT	Location	Length	Formula	ΔG°	Z	$\overline{Z}_{\rm C}$
			Glutaredoxin				
GLRX1	P25373	Cytoplasm	110	$C_{549}H_{886}N_{146}O_{170}S_4$	-4565	-5.8	-0.182
GLRX2	P17695	Mitochondrion	143	$C_{715}H_{1161}N_{181}O_{213}S_5$	-5617	0.1	-0.255
GLRX3	Q03835	Nucleus	285	$C_{1444}H_{2195}N_{371}O_{463}S_{10}$	-12031	-24.5	-0.094
GLRX4	P32642	Nucleus	244	$C_{1226}H_{1910}N_{316}O_{389}S_6$	-10276	-17.8	-0.140
GLRX5	Q02784	Mitochondrion	150	$C_{762}H_{1200}N_{196}O_{227}S_6$	-5841	-6.1	-0.192
			Thioredoxin				
TRX1	P22217	Cytoplasm	102	$C_{502}H_{785}N_{123}O_{150}S_5$	-3969	-3.1	-0.211
TRX2	P22803	Cytoplasm	103	$C_{497}H_{780}N_{122}O_{153}S_5$	-4056	-3.1	-0.197
TRXB1	P29509	Cytoplasm	318	$C_{1509}H_{2412}N_{402}O_{471}S_{12}$	-12330	-4.7	-0.159
TRX3	P25372	Mitochondrion	127	$C_{651}H_{1049}N_{167}O_{181}S_{10}$	-4617	4.9	-0.255
TRXB2	P38816	Mitochondrion	342	$C_{1640}H_{2615}N_{449}O_{501}S_{14}$	-12841	-1.5	-0.145

a. Amino acid compositions of subcellular isoforms of glutaredoxin (GLRX), thioredoxin (TRX) and thioredoxin reductase (TRXB) in *S. cerevisiae* were taken from the SWISS-PROT database [71] (accession numbers shown in the table). Chemical formulas of nonionized proteins, and calculated standard molal Gibbs energy of formation from the elements (ΔG°, in kcal mol^{-1}, at 25 °C and 1 bar) and net ionization state (Z) at $\mathrm{pH} = 7$ of charged proteins are listed. Average nominal oxidation state of carbon ($\overline{Z}_{\rm C}$) was calculated using Eqn. (12).

if the turnover rates of the proteins are not all the same or if, through evolution, the genes coding for the proteins undergo different non-synonymous mutations. Model systems consisting of interacting proteins are useful targets for assessing the potential for chemical reactivity, which might occur on evolutionary time scales longer than the physical interactions.

The purpose of this study is to quantify using a metastable equilibrium reference state the responses of populations of model proteins for different subcellular compartments of *S. cerevisiae* to gradients of oxidation-reduction potential. There are two major parts to this paper. In the first part, the reactions corresponding to intercompartmental interactions between isoforms (or homologs) of particular enzymes and between proteologs are quantified by calculating the oxygen fugacities for equal chemical activities of the reacting proteins or proteologs in metastable equilibrium. A ranking of relative metastabilities of the proteologs is discussed. Specific known interactions between compartments are considered in order to derive values of the oxygen fugacity within compartments that best metastabilize the corresponding proteologs relative to those of other compartments. Equal-activity values of the oxygen fugacity in the reactions are used to predict a sequence of formation of model proteologs in response to a temporal oxidation-reduction gradient.

In the second part of this paper, the relative abundances of model proteins in metastable equilibrium are calculated and compared with measured abundances. The range of protein abundances in a metastable equilibrium population often approaches that seen in experiments over a narrow window of oxygen fugacity. Positive and negative correlations between the calculated and experimental relative abundances are found in some cases. Local energy minimization and its opposition in the cellular demands for selectivity in protein formation are discussed as possible processes leading to the observed patterns.

Results and Discussion

Calculated metastability relations are described below for intercompartmental interactions between the model homologs and proteologs, and for intracompartmental interactions among the most abundant proteins in compartments or the reference model complexes. Experimental comparisons and discussion of their implications are integrated with these results.

Relative metastabilities of subcellular homologs of redoxins

The cytoplasmic, nuclear and mitochondrial homologs of glutaredoxin [72, 73, 74] and thioredoxin/thioredoxin reductase [75, 11] in yeast cells represent the first model systems for subcellular environments studied here. The names and chemical formulas of these proteins are listed in Table 1, together with some computed properties. The average nominal oxidation state of carbon ($\overline{Z}_{\rm C}$) is a function of the relative proportions of the elements in the chemical formula (see Methods). These values are provided just to get some initial bearing on the differences in compositions of the proteins. In Table 1 the proteins with the lowest values of $\overline{Z}_C$ are the mitochondrial homologs and those with the highest values of $\overline{Z}_C$ are the nuclear homologs.

Because the current objective is to describe the compositions of populations of proteins in terms of a variable like oxidation-reduction potential, a quantity such as $\overline{Z}_C$ is not sufficient; it has no explicitly derivable relation to intensive properties that can be measured. The forces acting on chemical transformations among proteins can, however, be assessed by first writing chemical reactions denoting their formation. An example of this procedure is given further below for a specific model system. The basic methods that apply there were used throughout this study. The standard molal Gibbs energies (ΔG°) and net charges of ionized proteins at $\rm pH = 7$ are listed in Table 1 so that the results described below can be reproduced at this $\rm pH$.

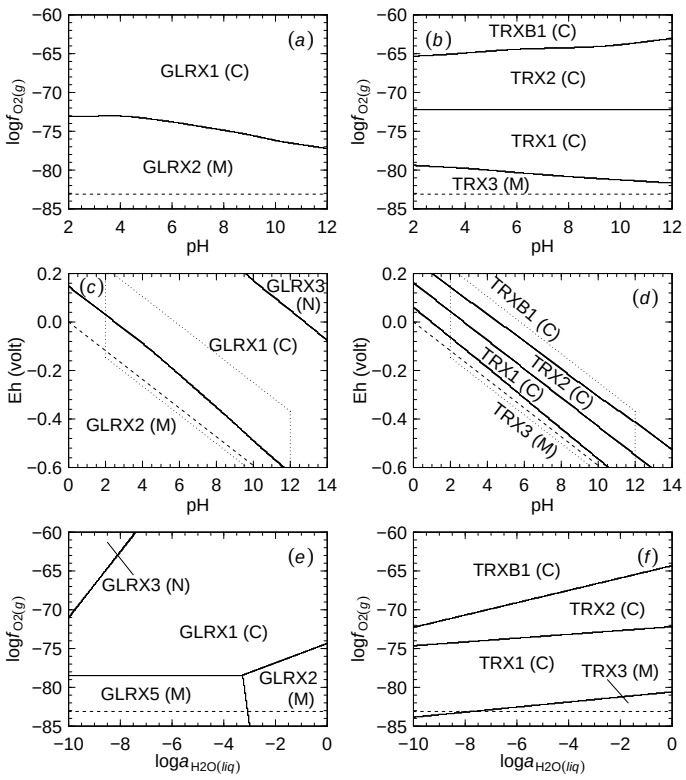

Figure 1: **Relative metastabilities of homologs of glutaredoxin and thioredoxin/thioredoxin reductase.** Predominance diagrams were generated for homologs of (*a*,*c*,*e*) glutaredoxin and of (*b*,*d*,*f*) thioredoxin/thioredoxin reductase in *S. cerevisiae*. The letters in parentheses following the labels indicate the subcellular compartment to which the protein is localized (C – cytoplasm; M – mitochondrion; N – nucleus). Calculations were performed for ionized proteins at 25 °C and 1 bar and for reference activities of basis species noted in the Methods. Reduction stability limits of $\rm H_2O$ are shown by dashed lines; the dotted lines in (*c*) and (*d*) correspond to the plot limits of (*a*) and (*b*).

In Figs. 1a and b the metastable equilibrium predominance limits of ionized proteins in the glutaredoxin and thioredoxin/thioredoxin reductase model systems are shown as a function of the logarithm of oxygen fugacity and $\rm pH$. Here, the predominant protein in a population is taken to be the one with the greatest chemical activity. The computation of the relative metastabilities of the proteins included all five model proteins in the glutaredoxin system as candidates, but note regarding Fig. 1a that only two of the five proteins appear on the diagram. Those that do not appear are less metastable, or have greater energy requirements for their formation over the range of conditions represented in Fig. 1a than either of the proteins appearing in the figure.

The equal-activity lines in these $\rm pH$ diagrams are curved because the ionization states of the proteins depend on $\rm pH$. The observation apparent in Fig. 1a that increasing $\log f_{\rm O_{2(g)}}$ favors formation of the cytoplasmic protein homolog relative to its mitochondrial counterpart is also true for the thioredoxin/thioredoxin reductase system shown in Fig. 1b. In comparing Figs. 1a and b note that in the latter figure, predominance fields for a greater number of candidate proteins appear, and that the predominance field boundary between mitochondrial and cytoplasmic proteins occurs at a lower oxidation-reduction potential. The dashed lines shown in each diagram of Fig. 1 are reference lines denoting the reduction stability limit of $\rm H_2O$ ($\log f_{\rm O_{2(g)}} \approx -83.1$ at 25 °C and 1 bar [76]).

Predominance diagrams as a function of Eh and pH for the glutaredoxin and thioredoxin/thioredoxin reductase systems are shown in Figs. 1c and d. Like $\log f_{\mathrm{O}_{2(g)}}$, Eh and pH together are a measure of the oxidation-reduction potential of the system; the different scales can be converted using Eqn. (11). The trapezoidal areas bounded by dotted lines in Figs. 1c and d show the ranges of Eh and pH corresponding to the $\log f_{\mathrm{O}_{2(g)}}$-$\mathrm{pH}$ diagrams of Figs. 1a and b. It can be deduced from these diagrams that if the upper $\log f_{\mathrm{O}_{2(g)}}$ limit of Fig. 1a were extended upward, this diagram would include a portion of the predominance field for the nuclear protein GLRX3.

It appears from Figs. 1a-b that increasing increasing $\log f_{\mathrm{O}_{2(g)}}$ at constant pH, or increasing pH at constant oxidation-reduction potential have similar consequences for the relative metastabilities of the cytoplasmic and mitochondrial homologs. In this analysis, however, pH does not appear to be a very descriptive variable; the magnitude of the effect of changing oxygen fugacity over several log units is greater than the effect of changing pH by several units. In further metastability calculations pH was set to 7. Also, because Eh itself is defined in terms of pH, the oxidation-reduction potential variable adopted below is $\log f_{\mathrm{O}_{2(g)}}$, which is more directly related to the potential of a thermodynamic component.

In Figs. 1e and f the logarithm of activity of water ($\log a_{\mathrm{H_2O}}$) appears as a variable. In Fig. 1e it can be seen that the formation of a nuclear homolog of GLRX is favored relative to the cytoplasmic homologs by decreasing activity of water and/or increasing oxygen fugacity, and that increasing relative metastabilities of the mitochondrial proteins are consistent with lower oxidation-reduction potentials and to some extent higher activities of water. In Fig. 1f it appears that the formation of the thioredoxin reductases relative to thioredoxins in each compartment is favored by increasing $f_{\mathrm{O}_{2(g)}}$, and that for the TRX the relative metastabilities of the mitochondrial proteins increase with decreasing $f_{\mathrm{O}_{2(g)}}$.

Comparison with subcellular redox measurements

Let us compare the positions of the predominance fields in Fig. 1 with measured subcellular redox states. The values of Eh derived from the concentrations of oxidized and reduced glutathione (GSSG and GSH, respectively) in extra- and subcellular environments reported in various studies [9, 77, 10, 79, 78] were converted to corresponding values of $\log f_{\mathrm{O}_{2(g)}}$ using Eqn. (11) in the Methods and are listed in Table 2. In order to fill in the table as completely as possible, it was necessary to consider measurements performed on eukaryotic cells other than those of *S. cerevisiae* (*e.g.*, HeLa [80] and mouse hybridoma [81] cells). The values of pH required for conversion of Eh to $\log f_{\mathrm{O}_{2(g)}}$ were also retrieved from the literature [83, 6, 7]. The computation of $\log f_{\mathrm{O}_{2(g)}}$ from Eh was performed at 25 °C and 1 bar and with $\log a_{\mathrm{H_2O}} = 0$. No measurements of vacuolar Eh have been reported, but it has been noted that Fe^{+3} predominates over Fe^{+2} in this compartment [85]. Hence, a nominal (and relatively very oxidizing) value of Eh for the vacuole was calculated that corresponds to equal activities of Fe^{+3} and Fe^{+2}.

Table 2: Nominal electrochemical characteristics of subcellular environments in eukaryotes. Values refer to yeast cells unless noted otherwise.

Environment	Eh, volt	pH	$\log f_{\mathrm{O}_{2(g)}}$ [m]
Extracellular (intestine)	-0.137 to -0.80 [a]	3 [g]	-83.3 to -79.4
Cytoplasm	-0.235 to -0.222 [b]	6.5 [h]	-75.9 to -75.0
Nucleus	– [c]	7.7 [i]	– [c]
Mitochondrion	-0.360 [d]	8 [j]	-78.3
Endoplasmic reticulum	-0.185 to -0.133 [e]	7.2 [k]	-69.7 to -66.2
Vacuole	$> +0.769$ [f]	6.2 [l]	> -9.2

a. [77] (*Homo sapiens*). **b.** The lower and upper values are taken from [78] and [79], respectively. **c.** The state of the GSSG/GSH couple in the nucleus is thought to be more reduced than in the cytoplasm [4]; see text. **d.** [10] (*Homo sapiens* HeLa [80] cells). **e.** [9] (*Mus musculus*: mouse hybridoma cells [81]). **f.** Calculated by combining the law of mass action for $\mathrm{Fe}^{+3} + e^- \rightleftharpoons \mathrm{Fe}^{+2}$ (standard molal Gibbs energies taken from [82]) with $a_{\mathrm{Fe}^{+3}} = a_{\mathrm{Fe}^{+2}}$ (see text). **g.** [83] (*Homo sapiens*). **h.** [6] (yeast). **i.** [84] (organism unspecified). **j.** [7] (HeLa) **k.** [8]. **l.** [5]. **m.** Values of Eh and pH listed here were combined with Eqn. (11) at $T = 25$ °C, $P = 1$ bar and $a_{\mathrm{H_2O}} = 1$ to generate the values of $\log f_{\mathrm{O}_{2(g)}}$.

The available measurements of redox states in compartments of eukaryotic cells can be summarized as, from most reducing to most oxidizing, mitochondria - nucleus - cytoplasm - endoplasmic reticulum - extracellular [4]. Strong redox gradients within the mitochondrion are essential to its function [86], which is not captured by the single values listed in Table 2. Comparison nevertheless with the computational results shown in Fig. 1 indicates that a relatively reducing environment does metastably favor the mitochondrial homolog.

Measurements of GSH/GSSG concentrations point to a lower redox state in the nucleus than in the cytoplasm,

Table 3: Overall protein compositions (proteologs) of compartments in yeast cells[a].

Location	Number	Length	Formula	ΔG°	Z	$\overline{Z}_{\mathrm{C}}$	$\log f_{\mathrm{O}_{2(g)}}$
actin	22	469.4	$C_{2316.7}H_{3636.4}N_{632.4}O_{721.5}S_{10}$	-18500	-5.2	-0.119	-75.0
ambiguous	123	821.3	$C_{4108.2}H_{6372.5}N_{1090.7}O_{1267.6}S_{30.3}$	-32181	-22.4	-0.123	NA
bud	57	429.9	$C_{2204.9}H_{3403.4}N_{574.6}O_{631.4}S_{19.7}$	-15314	6.8	-0.171	-75.2
bud.neck	11	905.2	$C_{4543.6}H_{7203.6}N_{1250.2}O_{1443.7}S_{26.6}$	-38092	-16.8	-0.113	-69.2
cell.periphery	38	826.1	$C_{4178.9}H_{6506.4}N_{1098.2}O_{1229.9}S_{33.8}$	-30639	1.8	-0.164	-74.7
cytoplasm	746	458.5	$C_{2294.5}H_{3623.3}N_{627.9}O_{704.8}S_{13.9}$	-18106	-3.9	-0.132	-74.6
early.Golgi	9	622.9	$C_{3197.7}H_{5067.7}N_{821.1}O_{972.8}S_{21.6}$	-25456	-19.9	-0.193	-75.3
endosome	30	457.6	$C_{2309.0}H_{3651.3}N_{626.1}O_{721.0}S_{13.9}$	-18833	-9.4	-0.131	-76.7
ER	197	309.7	$C_{1670.8}H_{2566.1}N_{415.5}O_{444}S_{10.9}$	-10292	9.4	-0.245	-75.5
ER.to.Golgi	5	594.9	$C_{2951.2}H_{4600.2}N_{790.8}O_{907.1}S_{18.9}$	-22855	-13.3	-0.127	NA
Golgi	14	478.2	$C_{2474.6}H_{3863.3}N_{642.7}O_{727.2}S_{15.1}$	-18229	-3.4	-0.182	-75.3
late.Golgi	29	602.4	$C_{3024.6}H_{4773.1}N_{802.9}O_{951.3}S_{17.9}$	-24962	-22.6	-0.141	-75.1
lipid.particle	17	502.1	$C_{2574.8}H_{3987.7}N_{673.3}O_{752}S_{17.5}$	-18691	-4.0	-0.167	-75.0
microtubule	10	497.0	$C_{2508.9}H_{3968.9}N_{689.9}O_{774.4}S_{17.9}$	-20024	-3.9	-0.125	-75.0
mitochondrion	426	484.9	$C_{2446.9}H_{3872.8}N_{669.4}O_{725.3}S_{16.0}$	-18244	6.8	-0.156	-75.9
nuclear.periphery	46	815.6	$C_{4110.6}H_{6516.1}N_{1091.9}O_{1272.1}S_{20.6}$	-33153	-11.5	-0.159	-75.2
nucleolus	60	564.3	$C_{2788.9}H_{4430.5}N_{771.5}O_{899.0}S_{13.3}$	-23928	-10.9	-0.104	-75.0
nucleus	453	572.1	$C_{2843.6}H_{4542.8}N_{802.2}O_{893.6}S_{20.3}$	-23525	-2.3	-0.108	-71.5
peroxisome	18	422.3	$C_{2117.9}H_{3334.8}N_{568.6}O_{642.0}S_{13.5}$	-16397	-2.0	-0.150	-74.8
punctate.composite	61	474.7	$C_{2355.7}H_{3719.7}N_{643.7}O_{763.3}S_{10.6}$	-20313	-22.6	-0.102	NA
spindle.pole	30	470.9	$C_{2391.1}H_{3818.9}N_{659.1}O_{749.4}S_{14.2}$	-19820	-6.3	-0.131	-78.8
vacuolar.membrane	45	762.2	$C_{3813.0}H_{5973.7}N_{1010.5}O_{1154.1}S_{26.2}$	-29221	-15.5	-0.153	-75.2
vacuole	67	511.7	$C_{2543.3}H_{3903.1}N_{650.7}O_{800.0}S_{16.8}$	-20239	-15.9	-0.125	-70.6

a. Chemical formulas of nonionized proteologs and standard molal Gibbs energy of formation from the elements (ΔG°, in kcal mol^{-1}, at 25 °C and 1 bar) and net ionization state (Z) at $\mathrm{pH} = 7$ of ionized proteologs were calculated using the overall amino acid compositions given in Table S1. Values of the nominal oxidation state of carbon ($\overline{Z}_{\mathrm{C}}$) were calculated using Eqn. (12). $\log f_{\mathrm{O}_{2(g)}}$ values for compartments were determined from the metastable equilibrium limits of subcellular interactions listed in Table 4.

but the chemical thermodynamic predictions show the nuclear proteins favored by relatively oxidizing conditions. Studies using nuclear magnetic resonance (NMR) showing that the hydration state of the nucleus is higher than the cytoplasm [16, 13] bring into question the prediction consistent with Fig. 1e that the formation of the nuclear proteins is favored relative to their cytoplasmic counterparts by decreasing activity of water. Also, mitochondrial pH is somewhat higher than that of the cytoplasm [6, 7], but in Figs. 1a and b it appears that the predicted energetic constraints favor the cytoplasmic proteins at higher pHs. These comparisons indicate that all metastable equilibrium constraints are not preserved in the spatial relationships of the homologous redoxins in the cell.

Relative metastabilities of proteologs

The chemical formulas and thermodynamic properties of the model proteologs – hypothetical proteins representing the overall amino acid compositions of compartments (see Methods) – are listed in Table 3. The predominance diagrams in Fig. 2 depicting the relative metastabilities of the model proteologs as a function of $\log f_{\mathrm{O}_{2(g)}}$ and $\log a_{\mathrm{H_2O}}$ were generated in sequential order. The first diagram in this figure corresponds to a system in which all 23 proteologs were considered. Subsequent diagrams in Fig. 2 were generated by eliminating from consideration some or all of the proteologs represented by predominance fields in the immediately preceding diagram. It can be seen in Fig. 2a that consideration of 23 proteologs resulted in predicted predominance fields for six proteins over the ranges of $\log f_{\mathrm{O}_{2(g)}}$ and $\log a_{\mathrm{H_2O}}$ shown in the diagram. Subsequent diagrams in the sequence represent proteologs with lower predicted relative metastabilities, i.e., higher energy requirements for formation relative to proteologs appearing earlier in the sequence.

There is a large difference between the relatively oxidized conditions of the endoplasmic reticulum reported in the literature (see Table 2) and the theoretically relatively reduced environment of the ER proteolog shown in

Fig. 2a. Also note the average nominal carbon oxidation state of the ER proteolog, which is the lowest of any in Table 3. A possible interpretation of these observations is that there is significant chemical heterogeneity within this compartment and a relatively high energy demand for the formation of these proteins in the oxidizing spaces. Nevertheless, the juxtaposition in the ER of very reduced proteins and high redox potential does permit a possible advantage: If the redox potential of the compartment were much lower, the proteins constituting the endoplasmic reticulum would become more favorable to produce than any other proteins (see below) ultimately localized to other compartments that are initially produced there. Perhaps in this way a high redox state could signal the production of cytoplasmic and secreted proteins and a drop in redox state the production of biosynthetic enzymes, i.e. the reproduction of the ER itself.

The proteologs appearing in successive diagrams in Fig. 2 are characterized by increasingly higher predicted energy requirements for their formation. Hence, the nuclear, cytoplasmic and mitochondrial proteologs appearing in Fig. 2c-d are relatively less metastable compared to those of actin, early Golgi and ER appearing in Fig. 2a. It is noteworthy that the proteologs representing the two cytoskeletal systems in yeast cells, actin and microtubule, appear at opposite ends of the energy spectrum. This prediction may be consistent with the observation that actin in different forms appears to be present at most stages of the cell cycle [87], but that the microtubule cytoskeleton grows during anaphase (i.e., the stage of the cell cycle characterized by physical separation of the chromosomes; [88]) and is degraded during other stages of the cell cycle [87, 88].

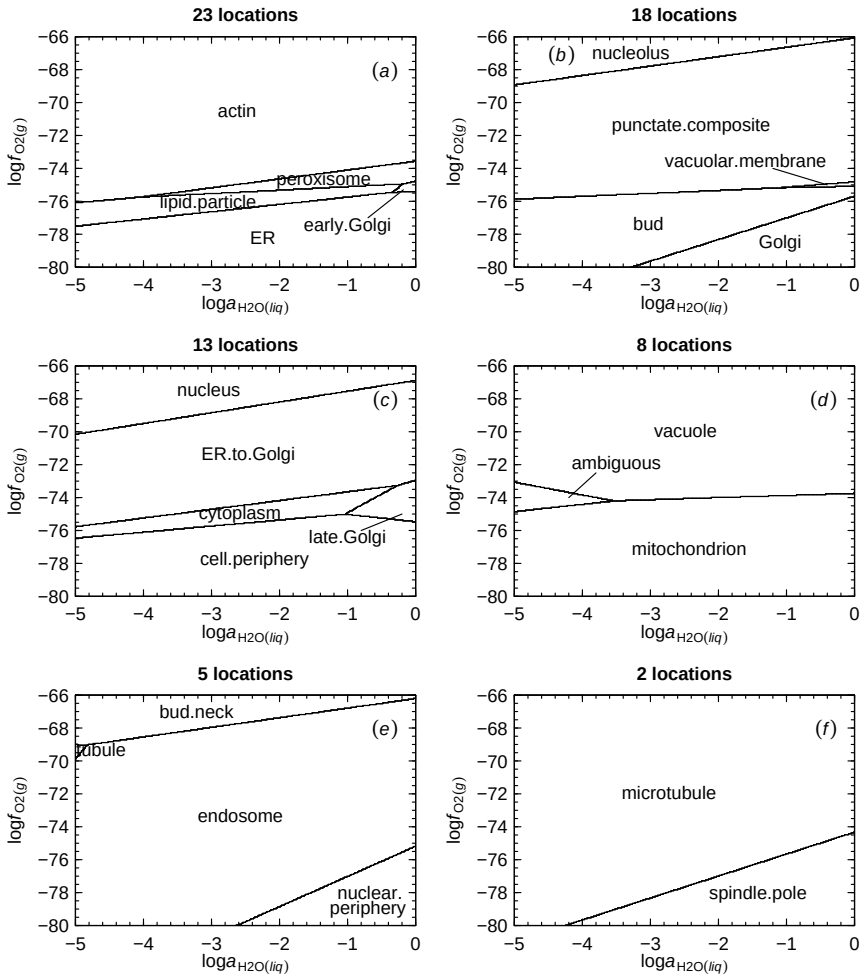

Figure 2: **Relative metastabilities of proteologs of compartments.** Predominance diagrams were generated as a function of $\log f_{\mathrm{O}_{2(g)}}$ and $\log a_{\mathrm{H_2O}}$ at 25 °C and 1 bar for the proteologs listed in Table 3. The diagram in (*a*) represents 23 model proteologs; diagrams in panels (*b*)–(*f*) represent successively fewer model proteologs.

The order of appearance of phases throughout a reaction sequence is determined by the relative stabilities of the phases [63]. Examples of the application of this notion in inorganic systems are the reaction series of metamorphic minerals, paragenetic sequences of mineralization [89], Ostwald ripening [90], and weathering reaction paths [91]. Can the relative metastabilities of proteins provide information about their order of appearance in the cell cycle?

The outcome of the mitotic cycle in *S. cerevisiae* is the growth of a new cell in the form of a bud [88]. Not all structures in the bud form simultaneously. Instead, it has been observed that [92] "the endoplasmic reticulum,

Table 4: Major intercompartmental protein interactions in yeast[a].

Interaction	Δn_{O_2}	$\log f_{O_{2(g)}}$	Interaction	Δn_{O_2}	$\log f_{O_{2(g)}}$
actin–bud	0.266	-75.1	**vacuole–bud**	0.223	-73.8
actin–bud.neck	0.078	-83.4	**vacuole–cell.periphery**	0.140	-73.4
actin–cell.periphery	0.183	-75.4	**vacuole–cytoplasm**	0.044	-70.7
actin–endosome	0.129	-76.6	**vacuole–endosome**	0.086	-74.1
actin–vacuolar.membrane	0.111	-75.1	**vacuole–late.Golgi**	0.072	-71.5
actin–mitochondrion	0.161	-75.8	nucleus–**actin**	-0.023	-88.7
actin–microtubule	0.124	-78.3	**nucleus**–microtubule	0.101	-75.9
microtubule–**bud**	0.142	-72.3	**nucleus–spindle.pole**	0.139	-75.5
microtubule–bud.neck	-0.045	-69.4	**nucleus–bud**	0.243	-73.8
microtubule–**cell.periphery**	0.059	-69.2	**nucleus**–bud.neck	0.056	-81.3
microtubule–**cytoplasm**	-0.037	-83.3	**nucleus–cytoplasm**	0.064	-71.7
microtubule–**spindle.pole**	0.038	-74.3	nucleus–**nucleolus**	-0.034	-78.7
spindle.pole–cytoplasm	-0.075	-78.7	**nuclear.periphery–bud.neck**	-0.080	-69.3
spindle.pole–**nuclear.periphery**	-0.004	-119.1	nuclear.periphery–**cytoplasm**	-0.072	-76.5
ER–cell.periphery	-0.460	-74.7	**nuclear.periphery–nucleus**	-0.136	-74.2
ER–cytoplasm	-0.557	-74.7	**nuclear.periphery–nucleolus**	-0.169	-75.1
ER–early.Golgi	-0.345	-75.4	**peroxisome**–cell.periphery	0.062	-78.7
ER–nuclear.periphery	-0.485	-74.4	**peroxisome**–cytoplasm	-0.034	-67.2
ER–peroxisome	-0.522	-75.2	**peroxisome–lipid.particle**	0.138	-74.9
Golgi–endosome	-0.199	-74.3	**peroxisome**–mitochondrion	0.040	-82.6
Golgi–vacuole	-0.285	-74.2	mitochondrion–**cell.periphery**	0.023	-72.0
Golgi–late.Golgi	-0.213	-75.2	**mitochondrion–cytoplasm**	-0.074	-75.4
Golgi–**early.Golgi**	-0.030	-84.4	**mitochondrion–nucleus**	-0.138	-73.7

a. Interactions between proteins in different subcellular locations in *S. cerevisiae* were identified in the literature. The calculated reaction coefficients on $O_{2(g)}$ and the metastable equilibrium value of $\log f_{O_{2(g)}}$ were calculated for each reaction between model proteologs. Names of locations shown in bold indicate that the model value of $\log f_{O_{2(g)}}$ for this compartment (Table 3) lies in the metastability range for the proteolog in the particular reaction.

Golgi, mitochondria, and vacuoles all begin to populate the bud well before anaphase and that their segregation into the bud does not require microtubules". The results in Fig. 2 indicate that the proteolog for bud is of comparable metastability relative to that of Golgi but it less metastable than the proteolog of ER. In the absence of energy input, it follows that there would be a chemical driving force to form the ER proteins at the expense of any of the bud that may be present. The appearance in the bud of the less-metastable mitochondrial proteins suggests that there is a source of energy to the bud that is nevertheless not sufficient to drive the formation of the proteins in the microtubule. The formation of these proteins may not be possible until the products of the mitochondrial reactions and other energy-rich metabolites have accumulated in the cell.

Intercompartmental protein interactions

The diagrams in Fig. 2 show the predominant metastability interactions between proteologs for different subcellular compartments. However, many subcellular interactions may in fact be meta-metastable with respect to Fig. 2. For example, interactions occur between proteins in the cytoplasm and nucleus [93], but the proteologs for these compartments do not share a reaction boundary in Fig. 2c. Below, known intercompartmental interactions are combined with the oxygen fugacity requirements for (meta-)metastable equilibrium of the proteologs to characterize compartmental oxidation-reduction potentials. These are used in the next section to explore a possible developmental reaction path.

To assess the biochemical evidence for specific interactions between proteins in different compartments in yeast cells, a series of review papers was surveyed [87, 94, 95, 93, 96, 97]. Statements implying interaction between proteins in different compartments were identified by scanning for action words including interact, are at, align, end at, organize, embed, move, associate, found, locate, extend, bisect, move, migrate, enter, attach, translocate, carry, sort, composed of, line, dock and fuse, recycle, transport, pinch, proceed, reach, degrade in, deliver, colocalize,

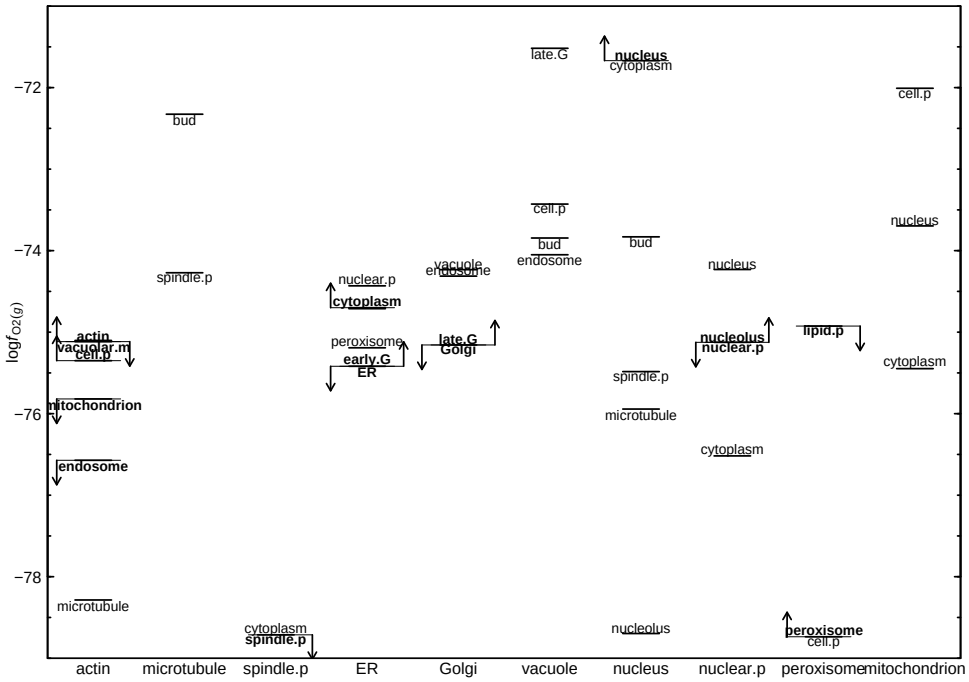

Figure 3: **Logarithms of oxygen fugacity for equal chemical activities of proteologs in intercompartmental interactions**. Metastable equilibrium values of $\log f_{\mathrm{O}_{2(g)}}$ were obtained for the model reactions listed in Table 4. Reactions are grouped by a common proteolog, listed along the bottom of the plot. Reactions that were used to derive model values of oxygen fugacity of compartments listed in Table 3 are denoted by arrows and bold lines and labels. The position of the reaction labels denotes the direction of the reaction that favors formation of the corresponding proteolog. The actin–bud and ER–cell periphery interactions were omitted from this plot to aid in clarity of labeling; they overlap with actin–vacuolar membrane and ER–cytoplasm, respectively.

contain, associate, separate, protrude, penetrate, cooperate, crosstalk, anchor, reside, continuous with, shuttle, oxidize, essential to, convey, arrange, import, and transcribe. The source statements are listed in Text S3 and simplified pairwise representations of the interactions are summarized in Table 4. Of 190 possible combinations between any two of the 20 subcellular compartments (this count excludes the ambiguous location and ER to Golgi and punctate composite, which did not appear in the literature survey), 46 interactions were identified through this survey.

Chemical reactions corresponding to each of the interactions listed in Table 4 were written between residue equivalents of the proteologs, with the reactant proteolog being the one on the left-hand side of the interaction and the product proteolog the one on the right-hand side. The reactions are listed in Table S2. Corresponding values of $\Delta n_{\mathrm{O}_{2(g)}}$ (reaction coefficient on $\mathrm{O}_{2(g)}$) are listed in Table 4 together with the values of $\log f_{\mathrm{O}_{2(g)}}$ where the calculated chemical activities of the two proteologs in each reaction are equal. Note that there are some reactions where the absolute value of $\Delta n_{\mathrm{O}_{2(g)}}$ is substantially smaller than the others; these include spindle pole–nuclear periphery, Golgi–early Golgi and nucleus–actin. Because of the small value of $\Delta n_{\mathrm{O}_{2(g)}}$ in these reactions, the values of $\log f_{\mathrm{O}_{2(g)}}$ for equal activities of these proteins tend to be more extreme than for other reactions. Note that the sign of $\Delta n_{\mathrm{O}_{2(g)}}$ denotes the thermodynamically favored direction of the reaction as $\log f_{\mathrm{O}_{2(g)}}$ is changed from its equal-activity value; for example, at $\log f_{\mathrm{O}_{2(g)}} = -75.1$, the proteologs of actin and bud metastably coexist with equal chemical activities, but at higher values that of actin predominates in metastable equilibrium.

The interactions listed in Table 4 were used to generate model values of the oxygen fugacity in each compartment that are listed in Table 3. The criterion used for this analysis was that the oxygen fugacity in a compartment should in as many cases as possible favor the formation of its proteolog relative to those of interacting compartments. For example, consider the proteolog for endosome, which occurs in three interactions listed in Table 4.

The endosomal proteolog is favored to form relative to that of actin by $\log f_{O_{2(g)}} < -76.6$ and relative to that of vacuole by $\log f_{O_{2(g)}} < -74.1$. In contrast, the endosomal proteolog is favored to form relative to the proteolog of Golgi by $\log f_{O_{2(g)}} > -74.3$. A single value of $\log f_{O_{2(g)}}$ can satisfy at most two of these constraints; the model value for endosome is taken to be just below the limit for its interaction with actin, or $\log f_{O_{2(g)}} = -76.7$ (Table 3). Because this value favors formation of the endosomal proteolog relative to those of actin and vacuole, the proteolog of endosome is listed in bold font in these interactions in Table 4, but is shown in normal font in the interaction with the Golgi proteolog. Similar reasoning was used to derive oxygen fugacities for the other subcellular compartments listed in Table 3, except for microtubule.

The outcome of the above analysis is summarized in Fig. 2, where the values of $\log f_{O_{2(g)}}$ for interactions that fall between -79 and -71 are plotted. The interactions are grouped by a common interacting proteolog so that differences between them can be more easily visualized. To avoid clutter, the reaction labels are generally restricted to the name of a single proteolog to indicate the direction of $\log f_{O_{2(g)}}$ change that favors its formation in the reaction. Model interactions that were used to constrain the limits of oxygen fugacities for one compartment (such as the actin–endosome interaction noted above) or two compartments (such as Golgi–late Golgi) are identified with one or two arrows, respectively, and the names of the corresponding proteologs are shown in bold font.

If the model compartmental values of $\log f_{O_{2(g)}}$ all favored formation of the corresponding proteologs relative to their interacting partners, the name of every proteolog would appear in bold font in Table 4. This is only the case, however, for some proteologs such as that of actin, where $\log f_{O_{2(g)}} - 75$ favors formation of this proteolog relative to any of its interacting partners. At the same oxygen fugacity, it can be shown that the proteolog for microtubule is unmetastable with respect to any of its interacting partners except for bud neck. Notably, the proteolog for microtubule only becomes relatively metastable at high oxygen fugacities (w.r.t. bud, cell periphery and spindle pole) or at low oxygen fugacities (w.r.t. actin, cytoplasm and nucleus). Hence, the value of $\log f_{O_{2(g)}} - 75$ taken here for the microtubule compartment is different from all the others, in that this represents conditions where the formation of its proteolog is more unfavorable than that of any of its interacting partners.

Sequential formation driven by oxygen fugacity gradients

We have already seen theoretical evidence that the microtubule is a relatively unmetastable assemblage of proteins in the cell. It is known in spite of this that the microtubule as well as the spindle pole are essential in cellular division [87]. Can the metastable equilibrium relationships reveal anything about the origins of the interactions of the microtubule and spindle pole in this process? The following thought experiment explores why the irreversible formation of proteologs might follow a sequence that is related to metastable equilibrium thermodynamic relationships.

To start, consider a permeable sac consisting of the cytoplasmic proteolog, which we will expose to a changing oxidation-reduction environment. The oxidation-reduction program will begin at $\log f_{O_{2(g)}} = -75$, drop to $\log f_{O_{2(g)}} = -83.5$, increase to $\log f_{O_{2(g)}} = -69$ and return to $\log f_{O_{2(g)}} = -75$. At any point along this program the only reactions we will consider are those involving the proteologs of microtubule or spindle pole. Let us assume in addition that none of these reactions proceeds to completion, and that any reaction may only proceed while $\log f_{O_{2(g)}}$ is near the equal-activity value for the reaction. Keeping in mind that no mechanism for the reactions is implied here, it may still be worthwhile to note that others have observed near-equilibrium concentrations of substrates in a subset of enzymatically catalyzed reactions [98, 99].

At $\log f_{O_{2(g)}} = -75$, no reaction occurs because the conditions coincide with the metastability field of the cytoplasmic proteolog relative to either microtubule or spindle pole. As soon as the $\log f_{O_{2(g)}}$ decreases below -78.7, some of the spindle pole proteolog may form irreversibly at the expense of the cytoplasmic proteolog. Below $\log f_{O_{2(g)}} = -83.3$, the microtubular proteolog can begin to form at the expense of the cytoplasmic proteolog. At $\log f_{O_{2(g)}} = -83.5$ both of these reactions may favorably proceed, and we begin now to increase $\log f_{O_{2(g)}}$. As we pass $\log f_{O_{2(g)}} = -83.3$, then $\log f_{O_{2(g)}} = -78.7$ going in the positive direction, some of the proteolog of microtubule, then spindle pole can react irreversibly to form the cytoplasmic proteolog. These are the opposite of the first two irreversible reactions.

As long as the current and following reactions do not proceed to completion, there will be a population of the microtubule and spindle pole proteologs available to react. Above $\log f_{O_{2(g)}} = -78.7$, where the formation of the cytoplasmic proteolog becomes favored relative to spindle pole (see above), the proteolog of actin may also favorably form at the expense of that of microtubule. The nuclear proteolog can form above $\log f_{O_{2(g)}} = -75.9$ at the expense of the microtubular proteolog, and above $\log f_{O_{2(g)}} = -75.5$ at the expense of the spindle pole

Table 5: Hypothetical oxygen fugacity cycle and sequence of reactions of proteologs.

$\log f_{O_{2(g)}}$	Reaction	$\log f_{O_{2(g)}}$	Reaction
-75.0	**Begin**	-74.3	spindle.pole→microtubule
-78.7	cytoplasm→spindle.pole	-69.4	microtubule→bud.neck
-83.3	cytoplasm→microtubule	-69.0	**Maximum point**
-83.5	**Minimum point**	-69.2	microtubule→cell.periphery
-83.3	microtubule→cytoplasm	-69.4	bud.neck→microtubule
-78.7	spindle.pole→cytoplasm	-72.3	microtubule→bud
-78.7	microtubule→actin	-74.3	microtubule→spindle.pole
-75.9	microtubule→nucleus	-75.0	**End**
-75.5	spindle.pole→nucleus		

proteolog. We now momentarily pass through our starting point, $\log f_{O_{2(g)}} = -75$. So far, the proteologs from spindle pole, microtubule, actin and nucleus, in that order, may have formed as a result of irreversible reactions of the original cytoplasmic proteolog. Also, the proteologs of microtubule and spindle pole may have been subsequently partially degraded after their possible formation.

Now, as $\log f_{O_{2(g)}}$ is increased above -74.3, the proteolog of spindle pole becomes unmetastable relative to that of microtubule. Above $\log f_{O_{2(g)}} = -69.4$, the proteolog of bud neck may be formed irreversibly at the expense of that of microtubule. At our maximum $\log f_{O_{2(g)}} = -69$ this reaction can continue, but as we drop below $\log f_{O_{2(g)}} = -69.2$ it may be joined by formation of the proteolog of cell periphery. Below $\log f_{O_{2(g)}} = -69.4$ any proteolog of bud neck that may have formed becomes unmetastable relative to that of microtubule. Below $\log f_{O_{2(g)}} = -72.3$ any proteolog of microtubule that remains may degrade in favor of formation of the proteolog of bud. Finally, as we drop past $\log f_{O_{2(g)}} = -74.3$ and return to our starting point of $\log f_{O_{2(g)}} = -75$ the proteolog of spindle pole once again becomes relatively metastable instead of microtubule. In summary, at $\log f_{O_{2(g)}} > -75$ the potential arises for formation of proteologs of the microtubule, bud neck, cell periphery, bud and spindle pole, as well as for retrograde reactions that may destroy the proteolog of microtubule.

It is important to emphasize the qualified nature of these predictions; all we know from thermodynamics is that any of these reactions could have progressed in the direction of a local Gibbs energy minimum. Whether and to what extent they actually move forward is a consequence of the reaction mechanism. The purpose of this analysis is not to suggest any mechanism but to ask whether work performed by control of $\log f_{O_{2(g)}}$ may energize such a mechanism. The enzymatic properties of the proteins themselves are probably essential in any actual mechanism. It is encouraging to observe that at and below the starting $\log f_{O_{2(g)}} = -75$ the proteolog of endoplasmic reticulum is favored to form relative to the cytoplasmic proteolog. Hence under these conditions there exists a potential for production of biosynthetic enzymes.

The results of this thought experiment are summarized in Table 5. The range of theoretical values of $f_{O_{2(g)}}$ required for the chemical transformations among the proteologs is between -83.5 and -69, which in terms of redox potential at 25 °C, 1 bar, $\mathrm{pH} = 7$ and $\log a_{H_2O} = 0$ correspond to $\mathrm{Eh} = -0.420$V and $\mathrm{Eh} = -0.205$V, respectively (Eqn. 11). The former value is just below the stability limit for water ($\log f_{O_{2(g)}} = -83.1$) but the redox state of the NADPH/NADP$^+$ pool in rat liver mitochondria might approach this value ($\mathrm{Eh} = -0.415$V [86]). The latter value is consistent with the state of human cells during differentiation ($\mathrm{Eh} = -0.200$V), which is about 0.040V higher than proliferating cells [100].

Oscillations in the redox state of yeast cells are coupled to many metabolic changes including protein transcription and turnover [101]. Reductive and oxidative phases in the metabolic cycle of yeast have been identified, with DNA replication occurring during the former and cell cycle initiation occurring at an advanced stage of the latter [102]. Oxidative stress was shown to hasten HeLa cells into anaphase by overcoming the normal spindle checkpoint mechanism [103]. Although the results shown in Table 5 do not directly address the synthesis of DNA, they do show that there is a potential for the formation of the nuclear proteolog during a relatively reducing part of the hypothetical $f_{O_{2(g)}}$ cycle. In the oxidizing part of this cycle, above $\log f_{O_{2(g)}} = -74.3$, the metastability of the proteolog for spindle pole is decreased, and at the highest oxidation-reduction potentials a favorable chemical potential field exists for metastable formation of the proteolog for bud neck. Hence, the notion that "a fundamental redox attractor underpins ... core cellular processes" [104] is in principle supported by the changing relative metastabilities of the proteologs as a function of oxidation-reduction potential.

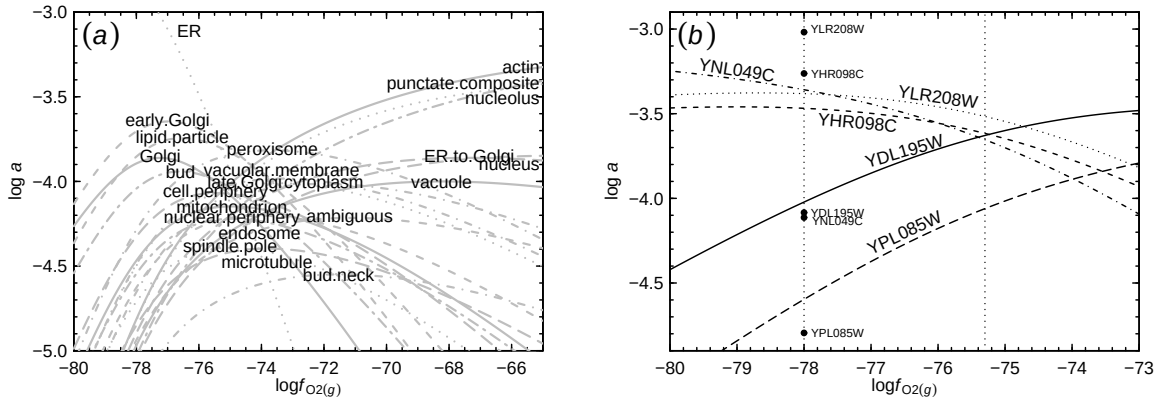

Figure 4: **Metastable equilibrium abundances of model proteologs and proteins as a function of oxygen fugacity.** Chemical speciation diagrams were generated as a function of $\log f_{\mathrm{O}_{2(g)}}$ at 25 °C and 1 bar and with total activity of protein residues equal to unity for (*a*) the proteologs shown in Table 1 and (*b*) the five proteins localized to ER to Golgi whose experimental abundances were reported in [105]. The rightmost dotted line in (*b*) indicates conditions where the calculated abundance ranking of the proteins is identical to that found in the experiments, and the leftmost dotted line where the calculated logarithms of activities have a lower overall deviation from experimental ones, which are indicated by the points. This value of $\log f_{\mathrm{O}_{2(g)}}$ (-78) was used to construct the corresponding diagram in Fig. 5.

Calculation of relative abundances of proteins

Above, the interactions between homologs (enzyme isoforms) in subcellular compartments and proteologs representing overall protein compositions in subcellular compartments were used to derive oxygen fugacity limits for metastable reaction of proteins in different compartments. In the second part of this study, attention is focused on the relative abundances and intracompartmental interactions of proteins.

The logarithms of activities of proteologs consistent with metastable equilibrium among all 23 model proteologs are plotted in Fig. 4a as a function of $\log f_{\mathrm{O}_{2(g)}}$. This diagram was generated based on metastable equilibrium among the residues of the proteins [70] in the same manner as described in detail below for a smaller set of proteins (those appearing in Fig. 4b). The purpose of Fig. 4a is to recapitulate the relationships shown in Fig. 2. Note that the same proteins predominate at the extremes of oxygen fugacity represented in 4a and in Fig. 1a (reducing – ER; oxidizing – actin) and that the proteolog of microtubule appears with low relative abundance. More importantly, perhaps, there is a minimum in the range of calculated activities of the proteologs around $\log f_{\mathrm{O}_{2(g)}} = -75$; changing oxidation-reduction potential alters not only the identity of the predominant protein in a metastably interacting population but also the relative abundances of all the others. There is probably not a single value of $\log f_{\mathrm{O}_{2(g)}}$ where the calculated relative abundances of the proteologs shown in Fig. 2 reflect the composition of the cell. Let us therefore look more closely at the relative abundances of proteins within compartments.

In Fig. 4b the relative abundances of the five model proteins localized exclusively to ER to Golgi are shown as a function of $\log f_{\mathrm{O}_{2(g)}}$. A worked-out example of the calculations leading to this figure, which method also underlies the generation of the other figures shown here, is presented in the following paragraphs.

The model proteins for ER to Golgi, in order of decreasing abundance in the cell reported by [105], are YLR208W, YHR098C, YDL195W, YNL049C and YPL085W. (For simplicity, the proteins are identified here by the names of the open reading frames (ORF).) The formula of the uncharged form of the first protein, YLR208W, is $\mathrm{C_{1485}H_{2274}N_{400}O_{449}S_4}$, and its amino acid sequence length is 297 residues. The standard molal Gibbs energy of formation from the elements (ΔG°) of this protein at 25 °C and 1 bar calculated using group additivity [69] is -10670 kcal mol^{-1} . At this temperature and pressure and at $\mathrm{pH} = 7$, group additivity can also be used [69] to calculate the charge of the protein (-10.8832) and the standard molal Gibbs energy of formation from the elements of the charged protein (-10880 kcal mol^{-1}). The formula of the protein in this ionization state is $\mathrm{C_{1485}H_{2263.1168}N_{400}O_{449}S_4^{-10.8832}}$. Dividing by the length of the protein, we find that the formula and standard molal Gibbs energy of formation from the elements of the residue equivalent of YLR208W are $\mathrm{C_{5.0000}H_{7.6199}N_{1.3468}O_{1.5118}S_{0.0135}^{-0.0366}}$ and -36.633 kcal mol^{-1}, respectively.

The formation from basis species of the residue equivalent of YLR208W is consistent with

$$5.0000\mathrm{CO}_{2(aq)} + 1.7946\mathrm{H_2O} + 1.3468\mathrm{NH}_{3(aq)} + 0.0135\mathrm{H_2S}_{(aq)} \rightleftharpoons \mathrm{C}_{5.0000}\mathrm{H}_{7.6199}\mathrm{N}_{1.3468}\mathrm{O}_{1.5118}\mathrm{S}_{0.0135}^{-0.0366} + 5.1414\mathrm{O}_{2(g)} + 0.0366\mathrm{H}^+ . \quad (1)$$

Similar reasoning can be applied to write the formation reaction of the residue equivalent of YHR098C as

$$4.9720\mathrm{CO}_{2(aq)} + 1.8708\mathrm{H_2O} + 1.3240\mathrm{NH}_{3(aq)} + 0.0441\mathrm{H_2S}_{(aq)} \rightleftharpoons \mathrm{C}_{4.9720}\mathrm{H}_{7.7882}\mathrm{N}_{1.3240}\mathrm{O}_{1.5231}\mathrm{S}_{0.0441}^{-0.0138} + 5.1464\mathrm{O}_{2(g)} + 0.0138\mathrm{H}^+ . \quad (2)$$

The double arrows signify that *a priori* one does not know the sign of the chemical affinity of either of these reactions.

At 929 residues, YHR098C is over 3 times as long as YLR208W, but in the formation reactions from the basis species of the residue equivalents of the two proteins, the coefficients on the basis species are similar. The difference between the coefficients of the same basis species in the reactions signifies the response (owing to moderation, i.e. LeChatelier's principle [106]) of the metastable equilibrium assemblage to changes in the corresponding chemical activity or fugacity. For example, because $\nu_{\mathrm{CO_2},1} < \nu_{\mathrm{CO_2},2}$, $\nu_{\mathrm{NH_3},1} < \nu_{\mathrm{NH_3},2}$ and $\nu_{\mathrm{O_2},1} < \nu_{\mathrm{O_2},2}$, increasing $a_{\mathrm{CO}_{2(aq)}}$, $a_{\mathrm{NH}_{3(aq)}}$ or $f_{\mathrm{O}_{2(g)}}$ at constant T, P and chemical activities of the other basis species shifts the metastable equilibrium in favor of YLR208W at the expense of YHR098C. Here, ν_i denotes the reaction coefficient of the ith basis species or protein, which is negative for reactants and positive for products as written. Conversely, because $\nu_{\mathrm{H_2O},1} > \nu_{\mathrm{H_2O},2}$, $\nu_{\mathrm{H_2S},1} > \nu_{\mathrm{H_2S},2}$ and $\nu_{\mathrm{H^+},1} > \nu_{\mathrm{H^+},2}$ increasing $a_{\mathrm{H_2O}}$, $a_{\mathrm{H_2S}_{(aq)}}$ or $a_{\mathrm{H^+}}$ (decreasing pH) at constant T, P and chemical activities of the other basis species shifts the metastable equilibrium in favor of YHR098C at the expense of YLR208W. The magnitude of the effect is proportional to the size of the difference between the coefficients of the basis species in the reactions, and it can be quantified for a specific model system using the following calculations.

To assess the relative abundances of the proteins in metastable equilibrium, we proceed by calculating the chemical affinities of each of the formation reactions. The chemical affinity ($\boldsymbol{A}$) is calculated by combining the equilibrium constant (K) with the reaction activity product (Q) according to [107]

$$\boldsymbol{A}/2.303RT = \log\left(K/Q\right) = \log\left(\frac{-\Delta G_r^\circ/2.303RT}{\prod a_i^{\nu_i}}\right), \quad (3)$$

where 2.303 is the natural logarithm of 10, R stands for the gas constant, T is temperature in degrees Kelvin, ΔG_r° is the standard molal Gibbs energy of the reaction, and a_i and ν_i represent the chemical activity and reaction coefficient of the ith basis species or species of interest (i.e., residue equivalent of the protein) in the reaction. Let us calculate ΔG_r° (in kcal mol^{-1}) of Reaction 1 by writing

$$\begin{aligned}\Delta G_1^\circ = {} & 1 \times -36.633 + 5.1414 \times 0 + 0.0366 \times 0 \\ & - 5.0000 \times -92.250 - 1.7946 \times -56.688 \\ & - 1.3468 \times -6.383 - 0.0135 \times -6.673 \\ & = 535.036 . \end{aligned} \quad (4)$$

In Eqn. (4) the values of ΔG° of $\mathrm{O}_{2(g)}$ and H^+ are both zero, which are consistent with the standard state conventions for gases and the hydrogen ion convention used in solution chemistry. The values of ΔG° of the other basis species are taken from the literature [108, 109, 110]. The value of $\log K_1$ consistent with Eqn. (4) is -392.19.

We now calculate the activity product of the reaction using

$$\begin{aligned}\log Q_1 = {} & 1 \times 0 + 5.1414 \times -75.3 + 0.0366 \times -7 \\ & - 5.0000 \times -3 - 1.7946 \times 0 - 1.3468 \times -4 - 0.0135 \times -7 \\ & = -366.92 . \end{aligned} \quad (5)$$

The values of a_i used to write Eqn. (5) are the reference values listed in the Methods for $a_{\mathrm{CO}_{2(aq)}}$, $a_{\mathrm{H_2O}}$, $a_{\mathrm{NH}_{3(aq)}}$, $a_{\mathrm{H_2S}_{(aq)}}$ and $a_{\mathrm{H^+}}$. The value of $f_{\mathrm{O}_{2(g)}}$ used in Eqn. (5) ($\log f_{\mathrm{O}_{2(g)}} = -75.3$) is also a reference value that, it will

be shown, characterizes a metastable equilibrium distribution of proteins that is rank-identical to the measured relative abundances of the proteins. Finally, the value of a of the residue equivalent of the protein in Eqn. (5) is set to a reference value of unity ($\log a = 0$). If we are only concerned with the relative abundances of the proteins in metastable equilibrium, the actual value used here does not matter so long as it is the same in the analogous calculations for the other proteins.

Combining Eqns. (3)–(5) yields $\boldsymbol{A}_1/2.303RT = -25.25$ (this is a non-dimensional number). Following the same procedure for the other four proteins (YHR098C, YDL195W, YNL049C and YPL085W) results in $\boldsymbol{A}/2.303RT$ equal to -24.86, -24.74, -24.93 and -24.94, respectively. Now let us turn to the relative abundances of the proteins in metastable equilibrium, which we compute using a Boltzmann distribution for the relative abundances of the residue equivalents:

$$\frac{a_i}{a_t} = \frac{e^{\boldsymbol{A}_i/RT}}{\sum_{i=1}^{n} e^{\boldsymbol{A}_i/RT}}\,, \tag{6}$$

where a_t denotes the total activity of residue equivalents in the system and n stands for the number of proteins in the system. Note regarding the left-hand side of Eqn. (6) that because we are taking activity coefficients of unity, the ratio a_i/a_t is equal to the ratio of concentrations, or proportionally numbers, of residue equivalents in the system. There is not a negative sign in front of $\boldsymbol{A}/RT$ in the exponents Eqn. (6) because the chemical affinity is the negative of Gibbs energy change of the reaction. Note in addition that the values of $\boldsymbol{A}/2.303RT$ given above must be multiplied by $\ln 10 = 2.303$ before being substituted in Eqn. (6). By taking $a_t = 1$, we can combine Eqn. (6) with $\boldsymbol{A}/RT$ of each of the formation reactions to calculate chemical activities of the residue equivalents of the proteins equal to 0.0905, 0.2248, 0.2994, 0.1944 and 0.1909, respectively. The lengths of the proteins are 297, 929, 1273, 876 and 2195, so the corresponding logarithms of activities of the proteins are e.g. $\log(0.0905/297) = -3.52$ for YLR208W, and -3.61, -3.63, -3.65 and -4.06 for the remaining proteins, respectively.

If one now iterates calculation of the chemical affinities of the residue formation reactions using the calculated metastable equilibrium logarithms of activities of the residue equivalents (instead of the starting reference value of $\log a = 0$), the resulting chemical affinities for each formation reaction will be all equal and generally non-zero. This property of metastable equilibrium was used in [70] to describe specific application of a method using a system of linear equations for finding the metastable equilibrium state without explicitly writing Eqn. (6).

The results of the calculation described above correspond to the dotted line at $\log f_{O_{2(g)}} = -75.3$ in Fig. 4b. At this oxygen fugacity, the ranks of abundance of the model proteins in metastable equilibrium are identical to the ranks of experimental abundances. The figure was generated in whole by carrying out this procedure for different reference values of $\log f_{O_{2(g)}}$. It can be seen in Fig. 4a that there is a narrow range on either side of $\log f_{O_{2(g)}} = -75.3$ (ca. ± 0.05) where the relative abundances of the proteins in metastable equilibrium occur in the same rank order. Beyond these limits, changing $f_{O_{2(g)}}$ drives the composition of the metastable equilibrium assemblage to other states that do not overlap as closely with the experimental rankings. The experimental abundances of the proteins reported by [105] are 21400, 12200, 1840, 1720 and 358, respectively, in relative units. These abundances were scaled to the same total activity of residues (unity) used in the calculations to generate the experimental relative abundances plotted at the dashed line in Fig. 4b at $\log f_{O_{2(g)}} = -78$. Under these conditions, the metastable equilibrium abundances of the proteins do not occur in exactly the same rank order as the experimental ones, but there is a greater overall correspondence with the experimental relative abundances.

Relative abundances of proteins within compartments

The procedure outlined above for calculating the relative abundances of model proteins in ER to Golgi was repeated for each of the other compartments identified in [22]. Up to 50 experimentally most abundant proteins were chosen to model each of the compartments. The relative abundances of the proteins were calculated at 0.5 $\log$ unit increments from $\log f_{O_{2(g)}} = -82$ to -70.5. Scatterplots of the experimental vs. calculated relative abundances for each set of proteins are shown in Figure S1. These comparisons were visually assessed to regress values of $\log f_{O_{2(g)}}$, listed in Table 6, that yield the best fit between calculated and experimental relative abundances. The resulting calculated relative abundances are listed together with the experimental ones in Table S3; the best-fit scatterplots for each set of model proteins are shown in Fig. 5

The retrieval of optimal values of $\log f_{O_{2(g)}}$ was aided by also calculating the root mean square deviation (RMSD) of logarithms of activities using Eqn. (13) and the Spearman rank correlation coefficient (ρ; Eqn. 14) between experimental and calculated logarithms of activities. The dotted lines in Fig. 5 were drawn at one RMSD

Table 6: Oxygen fugacities, deviations and correlation coefficients in comparisons of intracompartmental protein interactions[a].

Most abundant proteins					Model complexes				
Location	n	$\log f_{O_{2(g)}}$	RMSD	ρ	Complex	n	$\log f_{O_{2(g)}}$	RMSD	ρ
actin	22	-75.5	0.61	0.19	1	5	-80.0	0.35	0.90
ambiguous	50	-74.5	0.90	0.16	2	7	-78.0	0.58	0.57
bud	50	-74.5	0.85	0.02	3	5	-75.0	0.25	0.80
bud.neck	11	-75.5	0.73	0.02	4	6	-75.0	0.80	0.71
cell.periphery	38	-74.5	0.63	0.42	5	4	-74.5	0.29	0.80
cytoplasm	50	-77.0	1.20	0.16	6	7	-80.0	0.20	0.96
early.Golgi	9	-74.0	0.72	0.45	7	4	-75.5	0.21	0.80
endosome	30	-75.5	0.98	0.04	8	4	-73.5	0.60	-1.00
ER	49	-76.0	0.94	0.09	9	3	-77.0	0.16	-1.00
ER.to.Golgi	5	-78.0	0.40	0.40	10	4	-76.5	0.82	-0.80
Golgi	14	-75.5	0.80	0.08	11	10	-74.5	0.76	-0.28
late.Golgi	29	-74.5	0.61	0.18	12	5	-76.5	1.74	-0.30
lipid.particle	17	-78.0	0.92	0.23	13	12	-74.5	0.97	0.01
microtubule	10	-75.0	0.61	0.36	14	7	-73.0	1.05	-0.93
mitochondrion	50	-76.0	0.49	0.43	15	17	-74.0	0.49	-0.14
nuclear.periphery	46	-76.0	0.62	0.32	16	23	-76.0	0.43	0.53
nucleolus	50	-75.5	0.79	-0.18	17	6	-74.0	0.57	0.66
nucleus	50	-75.0	0.80	-0.02	18	5	-79.0	0.25	0.90
peroxisome	18	-75.5	0.55	0.56	19	8	-76.0	0.39	0.91
punctate.composite	49	-74.0	0.78	0.19	20	15	-74.5	0.59	0.66
spindle.pole	30	-76.0	1.07	-0.13	21	5	-80.0	1.06	0.60
vacuolar.membrane	45	-76.5	1.07	0.36	22	15	-78.5	1.14	0.14
vacuole	50	-74.5	1.49	-0.02	23	9	-78.0	0.93	0.32

a. Values of $\log f_{O_{2(g)}}$ in each location were regressed by comparing calculated and experimental logarithms of activities of the most abundant proteins in different subcellular locations and of selected complexes for each location (Figure S1). n denotes the number of model proteins used in the calculations. RMSD values were calculated using Eqn. (13), and ρ denotes the Spearman rank correlation coefficient, calculated using Eqn. (14).

on either side of the one-to-one correspondence, denoted by the solid lines in this figure. The RMSD values were used to identify outliers that are identified in Fig. 5 by letters and open symbols and that are listed in Table 7. To aid in distinguishing the points, they were assigned colors on a red-blue scale that denotes the average nominal oxidation state of carbon of the protein (Eqn. 12).

There is a considerable degree of scatter apparent in many of the plots shown in Fig. 5, so a low significance is attached with the $\log f_{O_{2(g)}}$ values regressed from these comparisons. In specific cases such as late Golgi and nuclear periphery a lower overall deviation is apparent and there is a visual indication of a positive correlation between the calculated and experimental relative abundances. Because they were regressed from individual noisy data, the values of $\log f_{O_{2(g)}}$ listed in Table 6 are probably not as representative of subcellular oxidation-reduction conditions as those listed in Table 3, which have the additional benefit of being partly based on known subcellular interactions (see above).

The comparisons depicted in Fig. 5 and in Figure S1 are important because they reveal that the range of protein abundance observed in cells is accessible in a metastable equilibrium assemblage at some values of $\log f_{O_{2(g)}}$. For example, the range of experimental abundances of the model proteins in actin covers about 1.6 orders of magnitude, while the calculated abundances vary over about 2.2 orders of magnitude. Extreme values of $\log f_{O_{2(g)}}$ tend to weaken this correspondence (Figure S1). The lowest degree of correspondence occurs for the cytoplasmic proteins, where ~ 6 orders of magnitude separate the predicted relative abundances of the top 50 most abundant proteins, which in the experiments have a dynamic range spanning about 1.2 orders of magnitude. The great degree of scatter apparent in many of the comparisons in Fig. 5a is troublesome. The scatter could be partly a consequence of including in the comparisons model proteins that do not actually interact with each other, despite their high relative abundances. To address this concern, a more selective approach was adopted below that takes account of fewer numbers of proteins that interact through the formation of complexes.

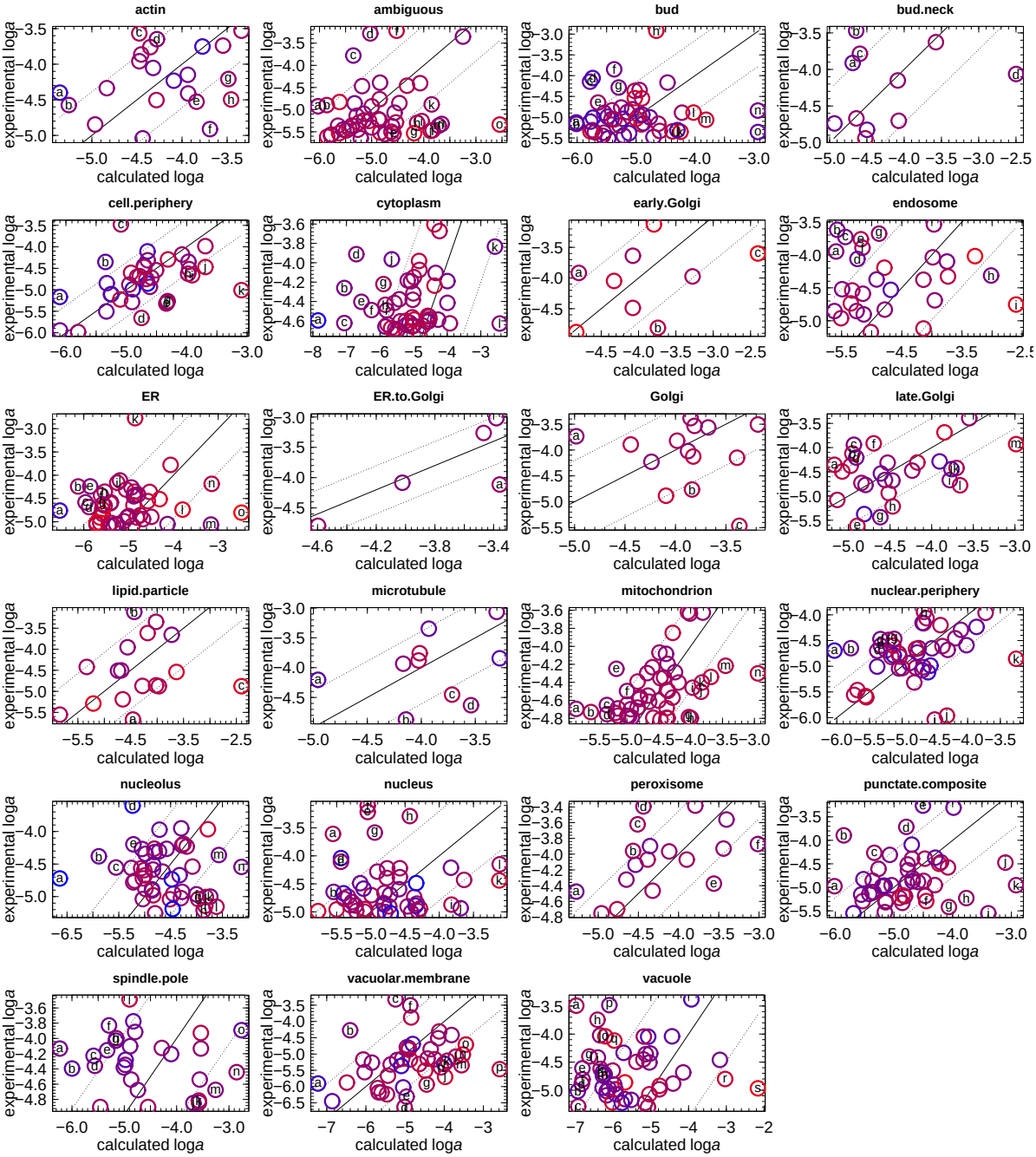

Figure 5: **Comparison of experimental and calculated logarithms of activities of proteins in compartments**. Red and blue colors denote, respectively, low and high average nominal carbon oxidation states ($\overline{Z}_{\mathrm{C}}$) of the protein. Dotted lines are positioned at one RMSD above and below one-to-one correspondence, which is denoted by the solid lines. Outlying points are labeled with letters that are keyed to the proteins in Table 7. The values of $\log f_{\mathrm{O}_{2(g)}}$ used in the calculations are listed in Table 6.

Table 7: Outliers in Fig. 5[a].

ID	ORF	ID	ORF	ID	ORF	ID	ORF	ID	ORF	ID	ORF
	actin		cell.periphery		endosome		microtubule		nucleolus		spindle.pole
a	YLR206W	a	YDR034W-B	a	YBR131W	a	YBL031W	a	YKR092C	a	YLR457C
b	YIL095W	b	YLL010C	b	YOR132W	b	YBL063W	b	YLL011W	b	YPL255W
c	YNR035C	c	YOR153W	c	YNR006W	c	YPL209C	c	YNL299W	c	YJR053W
d	YMR092C	d	YBR043C	d	YMR171C	d	YMR198W	d	YGR159C	d	YDR356W
e	YGR080W	e	YDR038C	e	YLR240W			e	YMR014W	e	YOR373W
f	YCR088W	f	YDR039C	f	YLR073C		mitochondrion	f	YJR063W	f	YGL061C
g	YIL062C	g	YDR040C	g	YGR206W	a	YIL125W	g	YGR271C-A	g	YKL089W
h	YOR367W	h	YHR146W	h	YKR035W-A	b	YNL063W	h	YCR086W	h	YIL144W
		i	YPR156C	i	YJR044C	c	YCL009C	i	YNR004W	i	YLR381W
	ambiguous	j	YLR413W			d	YHR051W	j	YLR367W	j	YPL233W
a	YGL021W	k	YOR094W		ER	e	YOR108W	k	YDR156W	k	YOL069W
b	YLR454W			a	YLR390W-A	f	YDR232W	l	YOR310C	l	YMR117C
c	YHR115C		cytoplasm	b	YEL002C	g	YMR083W	m	YLR221C	m	YDR016C
d	YER070W	a	YNL255C	c	YML012W	h	YKL040C	n	YNL113W	n	YDR201W
e	YIL065C	b	YBL027W	d	YJR131W	i	YFL018C			o	YKR083C
f	YJR011C	c	YBR084C-A	e	YDR221W	j	YPL078C		nucleus		
g	YPR139C	d	YER131W	f	YOR254C	k	YKL085W	a	YBR010W		vacuolar.membrane
h	YHR129C	e	YML026C	g	YNL258C	l	YDR298C	b	YNL251C	a	YPL180W
i	YHR025W	f	YDL082W	h	YML013W	m	YOR142W	c	YNR053C	b	YMR160W
j	YAR028W	g	YGL031C	i	YHR007C	n	YDL067C	d	YDR432W	c	YDL185W
k	YBR256C	h	YDR012W	j	YHR042W			e	YBR009C	d	YGL006W
l	YMR202W	i	YMR205C	k	YKL154W		nuclear.periphery	f	YNL030W	e	YLR447C
m	YJL034W	j	YPR035W	l	YDL128W	a	YDL088C	g	YLR153C	f	YBR127C
n	YMR214W	k	YDR382W	m	YKL096W-A	b	YGR202C	h	YBL002W	g	YBR207W
o	YJR085C	l	YOL039W	n	YBR106W	c	YKR095W	i	YDR190C	h	YML121W
				o	YEL027W	d	YAR002W	j	YIL021W	i	YDR486C
	bud		early.Golgi			e	YPR174C	k	YDR513W	j	YML018C
a	YDR309C	a	YGL223C		ER.to.Golgi	f	YER105C	l	YPL028W	k	YGR163W
b	YBL085W	b	YBL102W	a	YNL049C	g	YGL092W			l	YBR077C
c	YNL278W	c	YDR100W			h	YFR002W		punctate.composite	m	YOR332W
d	YNR049C				Golgi	i	YGL247W	a	YAR009C	n	YOL092W
e	YDR166C		lipid.particle	a	YDR245W	j	YLR450W	b	YGL200C	o	YOL129W
f	YPL032C	a	YCL005W	b	YNL041C	k	YHR133C	c	YJL186W	p	YHR039C-A
g	YER149C	b	YML008C	c	YLR268W			d	YNL243W		
h	YDR033W	c	YMR148W				peroxisome	e	YGR086C		vacuole
i	YGR191W				late.Golgi	a	YMR204C	f	YOL044W	a	YNL326C
j	YMR295C			a	YDR407C	b	YKL197C	g	YER071C	b	YER123W
k	YLR414C			b	YJL044C	c	YLR324W	h	YNL173C	c	YBR205W
l	YBR054W			c	YDR170C	d	YGL037C	i	YDR357C	d	YER001W
m	YLL028W			d	YBL010C	e	YDL022W	j	YBR052C	e	YDL211C
n	YPR124W			e	YMR218C	f	YGL153W	k	YDR032C	f	YOR099W
o	YOR304C-A			f	YGL083W					g	YPL019C
				g	YDR472W					h	YOL088C
	bud.neck			h	YPL259C					i	YBR199W
a	YJR092W			i	YBR254C					j	YDR483W
b	YPL116W			j	YLR330W					k	YIL005W
c	YHR023W			k	YKR068C					l	YLR300W
d	YPR188C			l	YKL135C					m	YPR159W
				m	YEL048C					n	YPL163C
										o	YJR161C
										p	YHR215W
										q	YNL336W
										r	YBR187W
										s	YGR105W

a. Proteins are listed whose calculated logarithm of activity differs from experimental values by more than the root mean square deviation shown in Table 6.

Relative abundances of proteins in complexes

The correspondence between the calculated and experimental relative abundances of the five model proteins in ER to Golgi raises the question of what characteristics of the proteins might be responsible for this result. Searching the functional annotations of these proteins reveals that they are part of the COPII coat complex [111]. The inclusion of the COPII complex above was largely unintentional, as the procedure there was to look at the most abundant proteins in given compartments. Nevertheless, the results for that model system suggested that focusing on specific complexes in other compartments could yield interesting results. Because the interactions of proteins to form complexes is essential in cellular structure and regulating functions of enzymes [51], factors that affect the relative abundances of the complexing proteins may be fundamental to the control of metabolic processes.

The model complexes used in this study are identified in Table 8. Each complex was nominally associated with a subcellular compartment based on the names and descriptions of the complexes available in the literature. Some exceptions are the cyclin-dependent protein kinase complex, the proteins of which are largely cytoplasmic and nuclear [22], but here is placed in the slot for the ambiguous location because no definitely ambiguously localized complexes could be identified. For a similar reason, the proteins listed in Table 8 under punctate composite are not part of a named complex but were chosen because they are localized to early Golgi in addition to the punctate

Table 8: Model proteins in complexes[a].

Name		ORF	
1. actin: Arp2/3 complex (423) ([112]; [114])			
Arc15		YIL062C	b
Arc18		YLR370C	
Arc19		YKL013C	
Arc35		YNR035C	a
Arc40		YBR234C	NA
Arp2	*	YDL029W	
Arp3		YJR065C	X
2. ambiguous: cyclin-dependent protein kinase complex (343)			
Cdc28	*	YBR160W	b
Cks1	*	YBR135W	a
Cln2	*	YPL256C	
Cys4	*	YGR155W	
Sic1	*	YLR079W	
Clb3	*	YDL155W	
Cln1	*	YMR199W	
3. bud: actin-associated motor protein complex 2 (49) [115]			
Myo2		YAL029C	
She4	*	YKL130C	b
Mlc1		YBR130C	
Myo1		YGL106W	X
Cmd1	*	YKL007W	
Myo5	*	YIL034C	a
4. bud.neck: septin complex (333) [117]			
Bud4		YJR092W	a
Cdc10		YCR002C	c
Cdc11		YJR076C	
Cdc12		YHR107C	
Cdc3		YLR314C	X
Shs1		YDL225W	b
Mdh1	*	YKL085W	
5. cell.periphery: exocyst complex (120)			
Exo84		YBR102C	NA
Sec10		YLR166C	
Sec3		YER008C	b
Sec5		YDR166C	a
Sec6		YIL068C	
Sec8		YPR055W	NA
6. cytoplasm: translation initiation factor eIF3 (45)			
Fun12		YAL035W	
Hcr1		YLR192C	c
Nip1		YMR309C	a
Prt1		YOR361C	
Rli1		YDR091C	
Rpg1		YBR079C	
Tif34		YMR146C	X
Tif35		YDR429C	NA
Tif5		YPR041W	b
7. early.Golgi: SNARE complex (113) [121]			
Dsl1	*	YNL258C	a
Sec39	*	YLR440C	
Tip20	*	YGL145W	
Ufe1	*	YOR075W	NA
Use1	*	YGL098W	
Pep12		YOR036W	X
Ykt6		YKL196C	X
8. endosome: ESCRT I & II complexes ([122]; [123])			
Vps23		YCL008C	X
Vps28	*	YPL065W	
Vps37		YLR119W	X
Mvb12		YGR206W	a
Vps22	*	YPL002C	b
Vps36	*	YLR417W	
Vps25		YJR102C	X
9. ER: signal recognition complex (52)			
Sec65		YML105C	NA
Srp14		YDL092W	
Srp54		YPR088C	X
Spp68	*	YPL243W	a
Srp72		YPL210C	b
10. ER.to.Golgi: coatomer COPII complex (340)			
Sec13		YLR208W	a
Sec16		YPL085W	
Sec23		YPR181C	X
Sfb2		YNL049C	
Sec24		YIL109C	NA
Grh1	*	YDR517W	
11. Golgi: Golgi transport complex (293)			
Cog1	*	YGL223C	
Cog2		YGR120C	b
Cog3		YER157W	
Cog4	*	YPR105C	
Cog5		YNL051W	c
Cog6		YNL041C	
Cog7		YGL005C	
Cog8	*	YML071C	
Iml1	*	YJR138W	
Nrp1	*	YDL167C	a
12. late.Golgi: retrograde protein complex (114) [118]			
Kar2	*	YJL034W	c
Vps52	*	YDR484W	
Vps53	*	YJL029C	NA
Vps54	*	YDR027C	a
Vps51	*	YKR020W	b
Scj1	*	YMR214W	
13. lipid.particle: sterol biosynthesis enzymes [119]			
Erg9	*	YHR190W	
Erg1	*	YGR175C	
Erg7		YHR072W	c
Erg11	*	YHR007C	
Erg24	*	YNL280C	
Erg25	*	YGR060W	
Erg26	*	YGL001C	
Erg27		YLR100W	NA
Erg6		YML008C	d
Erg2	*	YMR202W	
Erg3	*	YLR056W	a
Erg5	*	YMR015C	
Erg4	*	YGL012W	b
15. mitochondrion: mitochondrial ribosome small subunit (9)			
Ehd3		YDR036C	f
Mrp13		YGR084C	a
Mrp17		YKL003C	NA
Mrp21		YBL090W	h
Mrp4		YHL004W	
Mrp51		YPL118W	
Mrps16		YPL013C	NA
Mrps17		YMR188C	c
Mrps18		YNL306W	
Mrps28		YDR337W	e
Mrps5		YBR251W	d
Mrps8		YMR158W	X
Mrps9		YBR146W	X
Pet123		YOR158W	g
Rsm10		YDR041W	b
Rsm19		YNR037C	X
Rsm22		YKL155C	
Rsm23		YGL129C	
Rsm27		YGR215W	NA
Rsm7		YJR113C	
Mrp1		YDR347W	
Rsm25		YIL093C	
Nam9		YNL137C	
14. microtubule: DASH complex [113]			
Dam1		YGR113W	X
Duo1	*	YGL061C	a
Dad1	*	YDR016C	
Dad2	*	YKR083C	b
Spc19	*	YDR201W	
Spc34	*	YKR037C	NA
Ask1	*	YKL052C	
Dad3	*	YBR233W-A	
Dad4	*	YDR320C-A	
Hsk3		YKL138C-A	X
16. nuclear.periphery: nuclear pore complex [24]			
Nup60		YAR002W	
Nup170		YBL079W	
Asm4		YDL088C	a
Nup84		YDL116W	
Gle1		YDL207W	
Nup42		YDR192C	X
Nup157		YER105C	c
Gle2		YER107C	
Nic96		YFR002W	g
Nup145		YGL092W	
Seh1		YGL100W	X
Nup49		YGL172W	j
Nup57		YGR119C	
Nup159		YIL115C	
Nup192		YJL039C	
Nsp1		YJL041W	i
Nup82		YJL061W	f
Nup85		YJR042W	
Nup120		YKL057C	X
Nup100		YKL068W	h
Nup133		YKR082W	
Pom34		YLR018C	NA
Ndc1		YML031W	
Nup188		YML103C	e
Nup116	*	YMR047C	NA
Pom152		YMR129W	b
Nup53		YMR153W	
Nup1		YOR098C	d
Cdc31		YOR257W	X
17. nucleolus: small subunit processome (70) [120]			
Utp8		YGR128C	
Nan1		YPL126W	b
Utp10		YJL109C	a
Utp15		YMR093W	
Utp4		YDR324C	
Utp9		YHR196W	
18. nucleus: RNA polymerase I (30)			
Rpa49	*	YNL248C	NA
Rpa12	*	YJR063W	
Rpa190	*	YOR341W	
RPApa3	*	YOR340C	
Rpc40		YPR110C	a
Rpa135	*	YPR010C	
Rpb5		YBR154C	X
19. peroxisome: integral to peroxisomal membrane (GO:0005779)			
Ant1		YPR128C	
Inp2	*	YMR163C	
Pex12		YMR026C	
Pex15	*	YOL044W	
Pex22	*	YAL055W	b
Pex3		YDR329C	
Pex30		YLR324W	c
Pex31	*	YGR004W	a
Pex32	*	YBR168W	NA
Pxa1		YPL147W	X
Pxa2		YKL188C	X
20. punctate.composite: proteins localized here and early.Golgi			
Arl1		YBR164C	d
Apm3		YBR288C	
Bug1		YDL099W	
Arf1		YDL192W	
Luv1		YDR027C	a
Tvp23		YDR084C	
Dop1		YDR141C	a
Kei1		YDR367W	
Vrg4		YGL225W	
Apl6		YGR261C	
Aps3		YJL024C	c
Vps53		YJL029C	NA
Tvp38		YKR088C	
Ssp120		YLR250W	
NA		YMR010W	
NA		YMR253C	NA
Kex2		YNL238W	NA
Mon2		YNL297C	
21. spindle.pole: spindle-pole body complex (219) [116]			
Pfk1	*	YGR240C	a
Spc72		YAL047C	
Spc97		YHR172W	
Spc98		YNL126W	
Tub4		YLR212C	b
22. vacuolar.membrane: VO vacuolar ATPase complex (14)			
Emi2	*	YDR516C	
Vma6		YLR447C	
Vph2	*	YKL119C	a
Bni1	*	YNL271C	b
Drs2	*	YAL026C	
Gaa1	*	YLR088W	NA
Lys9	*	YNR050C	
Nop6	*	YDL213C	c
Pdc1	*	YLR044C	
Pgi1	*	YBR196C	
Vac8		YEL013W	
Vma10		YHR039C-A	d
Vma2		YBR127C	
Vma7		YGR020C	
Vph1		YOR270C	
Vtc4		YJL012C	X
Yor1	*	YGR281W	
Yra1	*	YDR381W	NA
23. vacuole: vacuolar proteases and other canonical proteins [12]			
Ape1	*	YKL103C	b
Ape3	*	YBR286W	
Lap3	*	YNL239W	
Pep4		YPL154C	NA
Prb1	*	YEL060C	
Prb1		YMR297W	
Ams1	*	YGL156W	a
Ath1		YPR026W	X
Pho8		YDR481C	
Vtc4		YJL012C	X
Ypt7	*	YML001W	c
Npc2		YDL046W	d
NA		YHR202W	NA

a. Numbers in parentheses refer to the ID of the complex, if available, from `http://yeast-complexes.embl.de` [124]. Compositions and localizations of complexes were also taken from references listed in square brackets. Symbols: "*" the protein was not localized in the compartment [22]; "X" or "NA" not tagged or no abundance [105]; "a", "b", etc. refer to outliers in Fig. 7.

composite characterization [22]. Other exceptions are the vacuolar model proteins (proteases and other canonical vacuolar proteins [12]), enzymes of the ergosterol biosynthetic pathway, some of which are associated with the lipid particle [119], and proteins integral to the peroxisomal membrane, which were identified using the Gene Ontology (GO) annotations in the SGD [111]. Where they could be found, the ID numbers of the complexes in a yeast complex database [124] are listed in parentheses in Table 8, as are literature references that describe the composition and/or localization of the complexes. If any of the proteins in the complexes do not localize [22] to the compartment shown in Table 8 they are marked with an asterisk; those proteins that were not present in the YeastGFP database or that are lacking an abundance count therein [105] are marked with "X" and "NA", respectively.

The calculated metastable equilibrium logarithms of activities of the proteins in each complex are shown as a function of $\log f_{O_{2(g)}}$ in Fig. 6. The calculated logarithms of activities of the proteins were compared with experimental ones by constructing scatterplots at $0.5 \log$ unit intervals from $\log f_{O_{2(g)}} = -82$ to -70.5, which are shown in Figure S1. As above, visual assessment of fit was the first resort to obtain values of $\log f_{O_{2(g)}}$ that maximize the correspondence with experimental relative abundances, but the RMSD and Spearman rank correlation coefficient were also considered in these comparisons. Because of the small sample size in many of the comparisons, the sign of the correlation coefficient is as useful as its magnitude in assessing the results. The resulting calculated relative abundances are listed together with the experimental ones in Table S4.

The number of model proteins in each of the complexes is less than the number of most abundant proteins in each compartment considered in the preceding section, so the visible decrease in scatter is expected. Some of the model complexes represented in Fig. 7 exhibit an apparent positive correlation between calculated and experimental logarithms of activities; these include translation initiation factor eIF3, nuclear pore complex and proteins integral to peroxisomal membrane. An inverse correlation between calculated and experimental logarithms of activities is apparent for proteins in the ESCRT I & II complexes, signal recognition complex, and DASH complex. A few of the other complexes (Golgi transport complex, sterol biosynthesis enzymes) exhibit very little overall correspondence between calculated and experimental logarithms of activities.

The results in Fig. 7 permit an interpretation of the relative energetic requirements for formation of different groups of interacting proteins. Take for example complex 14, which is the DASH complex that associates with the microtubule. An inverse correlation between the experimental and calculated relative abundances is apparent for this complex in Fig. 7. The RMSD between calculated and experimental logarithms of activities of proteins is 1.05, which is among the highest listed in Table 6. Note from Eqn. (3) that a ~ 1 log unit change in the chemical activity of a chemical species corresponds to a Gibbs energy difference equal to $2.303RT$. An average difference of ~ 1 between calculated and experimental logarithms of activity indicates that the formation of the proteins requires $2.303RT = 1364$ cal mol^{-1} per protein beyond what would be needed if the proteins formed in metastable equilibrium relative abundances. On the other hand, the formation in specific oxidation-reduction conditions of proteins making up translation initiation factor eIF3 and other assemblages where cellular abundances positively correlate with and span the same range as the metastable equilibrium distribution can proceed close to a local minimum energy required for protein formation.

Because of their relatively high energy demands, proteins in complexes such as the DASH complex and the spindle pole body are likely to be more dynamic in the cell. Although a positive rank correlation coefficient for the latter complex is reported in Table 6, at a higher oxygen fugacity ($\log f_{O_{2(g)}} = -76$) a strong inverse correlation obtains between experimental abundances and calculated metastable equilibrium relative abundances of the proteins in this complex (Figure S1). The finding made elsewhere of some inverse relationships between relative abundance of proteins and corresponding mRNA levels was also interpreted as evidence for additional effort on the part of the cell [125]. An inverse relationship that opposes equilibrium may be favored in evolution because of the strategic advantage of incorporating otherwise costly (rare) amino acids that increase enzymatic diversity [126]. The present results show that specific examples of inverse relationships in the relative abundances of proteins can be identified using a metastable equilibrium reference state that is conditioned by oxidation-reduction conditions. Chemical selectivity in the dynamic formation in the cell of high-energy proteins could lead to transient formation of complexes that function only under certain conditions. In contrast, complexing proteins that interact close to metastable equilibrium are more likely to be constitutively formed.

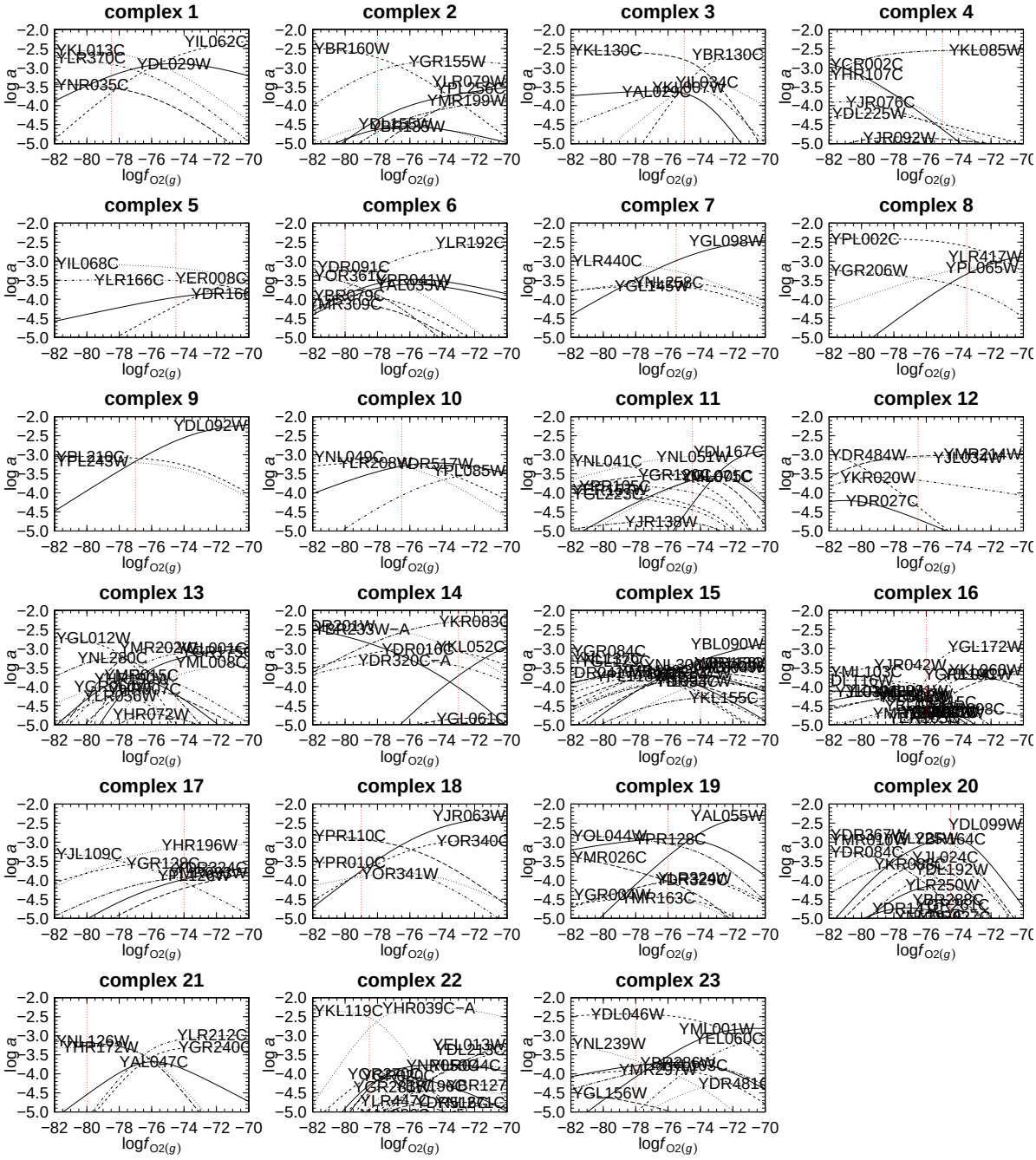

Figure 6: **Calculated logarithms of activities of model proteins in complexes**. The numbered complexes are identified in Table 8. Metastable equilibrium activities of proteins in the complexes were calculated as a function of $\log f_{O_{2(g)}}$ for total activity of residues set to unity. Dotted red lines denote values of $\log f_{O_{2(g)}}$ (listed in Table 6) and calculated relative abundances that were used in making Fig. 7.

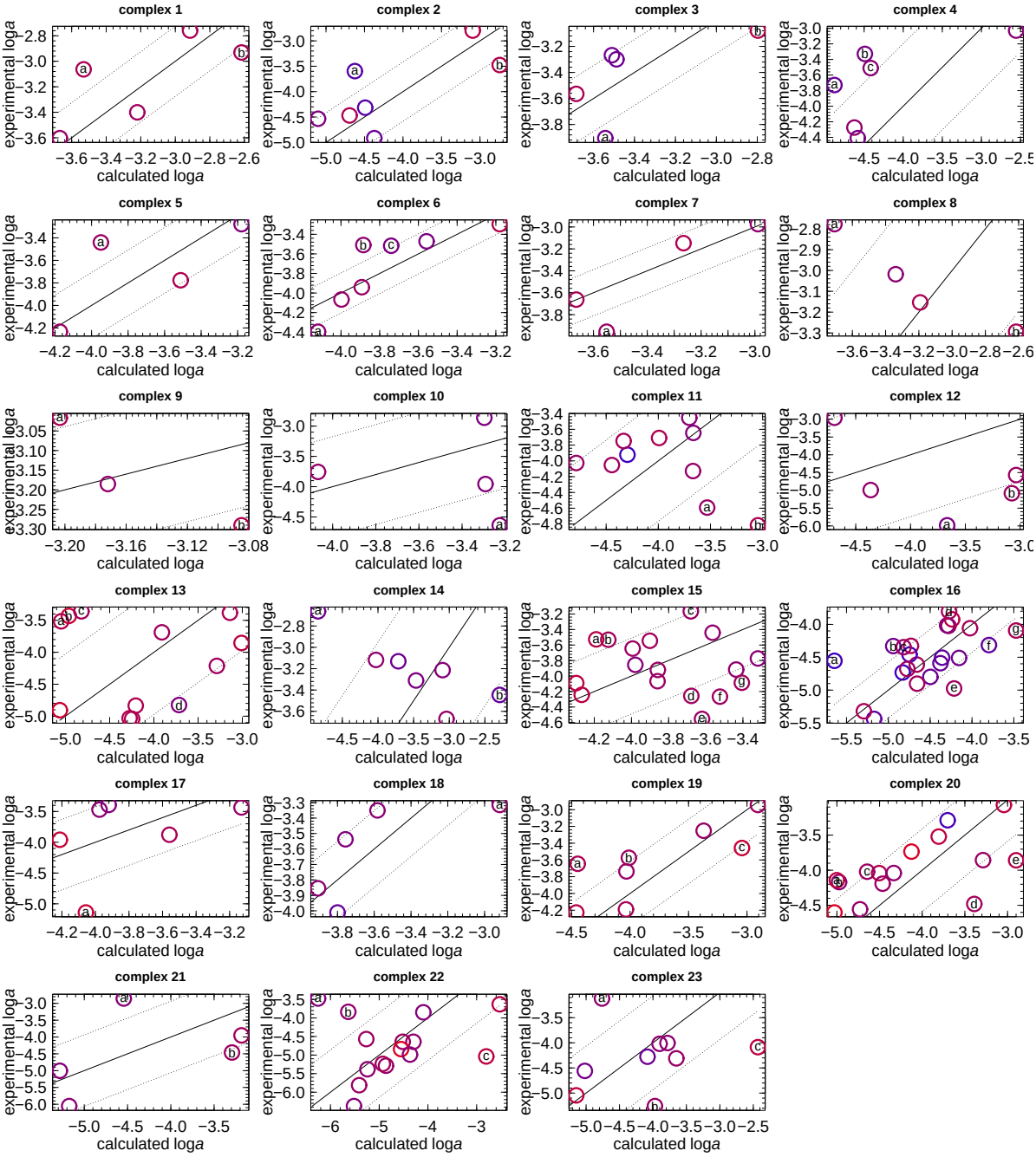

Figure 7: **Comparison of experimental and calculated logarithms of activities of interacting proteins**. Symbols are as in Fig. 5; the model proteins and the outliers are listed in Table 8.

Concluding Remarks

This study was concerned with thermodynamic selectivity of protein formation primarily as a function of one variable: oxidation-reduction potential represented by the logarithm of the fugacity of oxygen ($\log f_{O_{2(g)}}$). In reality, many variables are changing in cells, including the hydration state, pH, activity of CO_2 and H_2S, temperature and pressure. These all factor into the Gibbs energy changes accompanying the overall chemical transformation between proteins. Except for oxygen fugacity, the other variables were held constant in most of the calculations reported here. It is tempting to explore the effects of these variables on the compositions of metastable equilibrium assemblages. Incorporation into the framework of protein folding reactions and a non-ideality contribution, or excess Gibbs energy, that would encompass the effects of electrostatic interactions and macromolecular crowding is another target for expanding the scope of the thermodynamic characterizations.

The model results reported above were chosen in order to test specific predictions made using the hypothesis that the selection for or against metastable equilibrium has measurable consequences in organisms. The findings can be summarized as:

1. The oxidation-reduction potential ($\log f_{O_{2(g)}}$) limits of relative metastabilities of redoxin isoforms overlap with measured Eh (redox potential) in the cytoplasm and mitochondrion but not the nucleus.

2. The model proteologs represent the overall amino acid compositions of proteins in different compartments. At relatively low oxidation-reduction potential, proteologs in order of decreasing relative metastability are those of ER, Golgi, cell periphery, mitochondrion, nuclear periphery and spindle pole. At higher oxidation-reduction potential, proteologs in order of decreasing relative metastability are those of actin, nucleolus, nucleus, vacuole, bud neck and microtubule. At intermediate oxygen fugacities, proteologs of lipid particle, peroxisome and early Golgi are relatively metastable compared to those of cytoplasm, vacuolar membrane and late Golgi.

3. In a chemically reacting system starting with the cytoplasmic proteolog where all interactions include the proteologs of microtubule or spindle pole, environmental shifts in $\log f_{O_{2(g)}}$ going from -75 to -83.5 to -69 to -75 can drive the sequential formation of proteologs of spindle pole, microtubule, cytoplasm, actin, nucleus, cell periphery, bud neck and bud.

4. Oxidation-reduction potentials within $-78 < \log f_{O_{2(g)}} < -74$ give rise to metastable equilibrium populations of most abundant model proteins within compartments in which the range of protein abundance becomes closest to that seen in reported measurements. Substantial scatter is evident in the comparisons, but a moderate overall positive rank correlation was observed.

5. Closer fits between calculated and experimental relative abundances were obtained within $-80 < \log f_{O_{2(g)}} < -73$ by considering fewer numbers of model proteins that interact in complex formation. Strong positive correlations were found for, among others, cytoplasmic translation initiation factor eIF3 and nuclear pore complex; negative correlations were found for the microtubule-associated DASH complex and the endosomal ESCRT I & II complexes.

This study contributes to understanding the products of evolution by quantifying the extent of departure from metastable equilibrium in populations of interacting proteins. The observed positive correlations are consistent with a trend of some populations of interacting proteins to be imprinted with the consequences of local energy minimization in chemical reactions. These results and observations also support the notion that changing oxidation-reduction potential can selectively promote or hold back the reactions leading to formation of complexing proteins in relative abundances seen in the cell. Combining proteomic data with metastable equilibrium calculations is therefore a promising avenue for predicting complexes that form in specific oxidation-reduction conditions that vary temporally and spatially in biochemical systems.

Methods

The essential steps in the calculations reported here are 1) defining standard states, 2) identifying model proteins for systems of interest, 3) assessing the relative abundances of model proteins in metastable equilibrium, 4) visualizing the results of the calculations on speciation or predominance diagrams and 5) comparing the computational results with experimental biochemical and proteomic data.

Standard states and chemical activities

The activity of a species is fundamentally related to the chemical potential of the species by

$$\mu = \mu^\circ + RT \ln a \,, \tag{7}$$

where R and T represent, respectively, the gas constant and the temperature, μ and μ° stand for the chemical potential and standard chemical potential, respectively, and a denotes activity. No provision for activity coefficients of proteins or other species was used in this study; under this approximation, the activity of an aqueous species is equal to its concentration (molality).

The standard state for aqueous species including proteins specifies unit activity of the aqueous species in hypothetical one molal solution referenced to infinite dilution. The standard molal Gibbs energies of the proteins were calculated with the CHNOSZ software package [70] using group additivity properties and parameters taken from [69].

Proteologs: overall compositions of proteins in compartments

The overall amino acid compositions of proteins in 23 subcellular locations in *S. cerevisiae* were calculated by combining localization [22] and abundance [105] data for proteins measured in the YeastGFP project with amino acid compositions of proteins downloaded from the *Saccharomyces* Genome Database (SGD) [111]. Of 4155 ORF names listed in the YeastGFP dataset, all but 12 are present in SGD (the missing ones are YAR044W, YBR100W, YDR474C, YFL006W, YFR024C, YGL046W, YGR272C, YJL012C-A, YJL017W, YJL018W, YJL021C and YPR090W).

To generate proteologs that are most representative of each compartment, proteins that were annotated in the YeastGFP study as being localized to more than one compartment were excluded from this analysis (except for bud; see below), as were those for which no abundance was reported. The names of the open reading frames (ORFs) corresponding to the proteins in the YeastGFP data set were matched against the SGD's `protein_properties.tab` file downloaded on 2008-08-04. This search yielded a number of model proteins for each compartment, ranging from 5 (ER to Golgi) to 746 (cytoplasm); see Table 3. The names of the compartments used throughout the tables and figures in this paper correspond to the notation used in the YeastGFP data files (where spaces are replaced with a period).

It was found that no proteins with reported abundances and localized to the bud were exclusive to that compartment, hence all of the proteins localized there (which also have localizations in other compartments) were taken as models for the bud proteolog. The amino acid composition of the proteolog for each compartment was calculated by taking the sum of the compositions of each model protein for a compartment in proportion to its fractional abundance in the total model protein population of the compartment. The resulting amino acid compositions are listed in Table S1. The corresponding chemical formulas of the nonionized proteologs and the calculated standard molal Gibbs energies of formation from the elements at 25 °C and 1 bar of the ionized proteologs are shown in Table 3.

Metastability calculations

Diagrams showing the predominant proteins and the relative abundances of proteins in metastable equilibrium were generated using the CHNOSZ software package [70]. These calculations take account of formation reactions of the proteins written for their residue equivalents [70]. This approach is demonstrated in the Results for a specific model system.

The basis species, or perfectly mobile components of an open system [61], appearing in the formation reactions studied here are $CO_{2(aq)}$, H_2O, $NH_{3(aq)}$, $O_{2(g)}$, $H_2S_{(aq)}$ and H^+. The reference activities used for the basis species were 10^{-3}, 10^0, 10^{-4}, 10^{-7} and 10^{-7}, respectively, for $CO_{2(aq)}$, H_2O, $NH_{3(aq)}$, $H_2S_{(aq)}$ and H^+. In the case of diagrams showing Eh as a variable, the aqueous electron (e^-) was substituted for $O_{2(g)}$ in the basis species. Reference values for a_{e^-} or $f_{O_{2(g)}}$ are not listed here because one or the other is used as an independent variable in each of the calculations described above.

Conversion between scales of oxidation-reduction potential

Conversion between the $\log f_{O_{2(g)}}$ and Eh scales of oxidation-reduction potential can be made by first writing the half-cell reaction for the dissociation of H_2O as

$$H_2O \rightleftharpoons \frac{1}{2}O_{2(g)} + 2H^+ + 2e^- \,. \tag{8}$$

Taking $\mathrm{pH} = -\log a_{H^+}$ and $\mathrm{pe} = -\log a_{e^-}$, the logarithmic analog of the law of mass action for Reaction 8 can be written as

$$\log K_8 = \frac{1}{2}\log f_{O_{2(g)}} - 2\mathrm{pH} - 2\mathrm{pe} - \log a_{H_2O} \,, \tag{9}$$

where $\log K_8$ stands for the logarithm of the equilibrium constant of Reaction 8 as a function of temperature and pressure. Eh is related to pe by [127]

$$\mathrm{pe} = \frac{F}{2.303RT}\mathrm{Eh} \,, \tag{10}$$

where F and R denote the Faraday constant and the gas constant, respectively. Combining Eqns. (9) and (10) yields the following expression for Eh as a function of $\log f_{O_{2(g)}}$ and other variables:

$$\mathrm{Eh} = \frac{2.303RT}{F}\left(\frac{1}{2}\log f_{O_{2(g)}} - 2\mathrm{pH} - \log a_{H_2O} - \log K_8\right) \,. \tag{11}$$

At 25°C and 1 bar, $F/2.303RT = 16.903$ volt^{-1} and $\log K_8 = -41.55$; for $\mathrm{pH} = 7$ and $\log a_{H_2O} = 0$, a value of $\mathrm{Eh} = 0$ V corresponds to $\log f_{O_{2(g)}} = -55$. Eqn. (11) permits the conversion between Eh and $\log f_{O_{2(g)}}$ as well at other temperatures, pHs, and activities of H_2O.

Average nominal oxidation state of carbon

Let us write the chemical formula of a species of interest as $C_{n_C}H_{n_H}N_{n_N}O_{n_O}S^Z_{n_S}$, where Z denotes the net charge. The average nominal oxidation state of carbon ($\overline{Z}_C$) of this species is given by

$$\overline{Z}_C = \frac{Z - n_H + 2\,(n_O + n_S) + 3n_N}{n_C} \,. \tag{12}$$

Eqn. (12) is consistent with the electronegativity rules described in [128] and is compatible with the equation for average oxidation number of carbon used in [129]. For example, Eqn. (12) can be used to calculate the average nominal oxidation states of carbon in CO_2 and CH_4, which are $+4$ and -4, respectively. Note that the proportions of oxygen and other covalently-bonded heteroatoms contribute to the value of $\overline{Z}_C$ of a protein or other molecule, but that proton ionization does not alter the nominal carbon oxidation state, because of the opposite contributions from Z and n_H in Eqn. (12). In the 4143 proteins identified in the YeastGFP subcellular localization study and found in the *Saccharomyces* Genome Database, the minimum and maximum of $\overline{Z}_C$ are -0.414 and 0.390, respectively. Of the proteins in this dataset, six have $\overline{Z}_C < -0.35$ (YDR193W, YDR276C, YEL017C-A, YJL097W, YML007C-A, YMR292W) and six have $\overline{Z}_C > 0.15$ (YCL028W, YHR053C, YHR055C, YKR092C, YMR173W, YPL223C). The points in the scatterplots in this paper (Figs. 5 and 7 and Figure S1) are colored on a continuous red-blue scale according to the value of $\overline{Z}_C$ of the proteins, where maximum red occurs at $\overline{Z}_C = -0.35$ and maximum blue occurs at $\overline{Z}_C = 0.15$.

Comparison with experimental relative abundances

In comparison, experimental abundances of proteins in each model system were scaled so that the total chemical activity of residues was equal to unity.

The root mean square deviation between calculated and experimental logarithms of activities was calculated using

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{n} \left(X_{\mathrm{calc},i} - X_{\mathrm{expt},i}\right)^2}{n}}\,, \tag{13}$$

where $X_{\mathrm{calc},i}$ and $X_{\mathrm{expt},i}$ denote the calculated and experimental logarithms of activities and n stands for the number of proteins.x

The Spearman rank correlation coefficient (ρ) was calculated using

$$\rho = 1 - \frac{6d}{n\left(n^2 - 1\right)}\,, \tag{14}$$

where $d = \sum_{i=1}^{n} \left(x_{\mathrm{calc},i} - x_{\mathrm{expt},i}\right)^2$ and $x_{\mathrm{calc},i}$ and $x_{\mathrm{expt},i}$ stand for the ranks of the corresponding logarithms of activities.

Supporting Information

Figure S1: Comparisons of relative abundances of proteins (PDF)

Scatterplots of experimental vs. calculated abundance ranking and logarithm of activity of most abundant proteins and selected complexes in subcellular compartments are shown as a function of oxygen fugacity.

Table S1: Amino acid compositions of model proteologs (CSV)

Overall amino acid compositions of proteins in subcellular locations of *S. cerevisiae* were calculated from YeastGFP localization [22] and abundance [105] data downloaded from `http://yeastgfp.ucsd.edu/` combined with protein compositions downloaded from the *Saccharomyces* Genome Database (`http://www.yeastgenome.org/`). The amino acid compositions of the proteologs were used to calculate the properties listed in Table 3.

Table S2: Intercompartmental protein reactions (TXT)

This table lists chemical reactions between residue equivalents of proteologs for interacting compartments. The charges of the proteologs were calculated at 25 °C, 1 bar and $\mathrm{pH} = 7$.

Table S3: Abundance data for model proteins in compartments (CSV)

For the up to 50 most abundant model proteins in each compartment are listed the ORF name, sequence length, average nominal oxidation state of carbon (Eqn. 12), computed standard molal Gibbs energy at 25 °C and 1 bar of the ionized protein and charge at $\mathrm{pH} = 7$ and calculated and experimental logarithm of activity.

Table S4: Abundance data for protein complexes (CSV)

For the model complexes in each compartment (see Table 8) are listed the same properties as in Table S3.

Text S1: CHNOSZ software package (GZ)

This is the complete package (source code, documentation and data files) for the CHNOSZ program, which was used together with the program script (below) to perform the calculations in this study. The package is designed to be used with the R software environment `http://www.R-project.org`. Additional information about CHNOSZ is available in [70] and at `http://www.chnosz.net`.

Text S2: Program script and data files for generating figures (GZ)

This program script and supporting files were used to generate the figures shown above. It includes the script itself (plot.R), protein compositions (generated from the protein_properties.tab file downloaded from the *Saccharomyces* Genome Database), calculated standard molal thermodynamic properties of the proteins (to speed up calculations), YeastGFP protein localization and abundance data [22, 105], and a .csv version of Table 6. To generate the figures, the contents of the zip file should all be placed into the R working directory before loading CHNOSZ. Then read in the script with `source('plot.R')`. More details on the operation are provided at the top of the script file.

Text S3: Interactions between subcellular compartments in yeast (PDF)

This file lists statements from [87, 94, 95, 93, 96, 97] used to identify the interactions between proteins in different compartments of *Saccharomyces cerevisiae* that are listed in Table 4.

References

[1] Al-Habori M (1995) Microcompartmentation, metabolic channelling and carbohydrate metabolism. Int J Biochem Cell Biol 27:123 – 132.

[2] Aw TY (2000) Intracellular compartmentation of organelles and gradients of low molecular weight species. In: Walter H, Brooks DE, Srere PA, editors, Microcompartmentation and Phase Separation in Cytoplasm, San Diego: Academic Press, volume 192 of *Int. Rev. Cytol.* pp. 223 – 253.

[3] Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans R Soc B-Biol Sci 358:59 – 83.

[4] Hansen JM, Go YM, Jones DP (2006) Nuclear and mitochondrial compartmentation of oxidative stress and redox signaling. Annu Rev Pharmacol Toxicol 46:215 – 234.

[5] Preston RA, Murphy RF, Jones EW (1989) Assay of vacuolar pH in yeast and identification of acidification-defective mutants. Proc Natl Acad Sci U S A 86:7027–7031.

[6] Imai T, Ohno T (1995) Measurement of yeast intracellular pH by image processing and the change it undergoes during growth phase. J Biotech 38:165 – 172.

[7] Llopis J, McCaffery JM, Miyawaki A, Farquhar MG, Tsien RY (1998) Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc Natl Acad Sci U S A 95:6803 – 6808.

[8] Wu MM, Llopis J, Adams S, McCaffery JM, Kulomaa MS, et al. (2000) Organelle pH studies using targeted avidin and fluorescein-biotin. Chem Biol 7:197 – 209.

[9] Hwang C, Sinskey AJ, Lodish HF (1992) Oxidized redox state of glutathione in the endoplasmic reticulum. Science 257:1496 – 1502.

[10] Hanson GT, Aggeler R, Oglesbee D, Cannon M, Capaldi RA, et al. (2004) Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. J Biol Chem 279:13044 – 13053.

[11] Trotter EW, Grant CM (2005) Overlapping roles of the cytoplasmic and mitochondrial redox regulatory systems in the yeast *Saccharomyces cerevisiae*. Eukaryot Cell 4:392 – 400.

[12] Sarry JE, Chen S, Collum RP, Liang S, Peng M, et al. (2007) Analysis of the vacuolar luminal proteome of *Saccharomyces cerevisiae*. FEBS J 274:4287–4305.

[13] Morrill GA, Kostellow AB, Osterlow K, Gupta RK (1996) Differences in hydration state of nucleus and cytoplasm of the amphibian oocyte. J Membrane Biol 153:45 – 51.

[14] Garlid KD (2000) The state of water in biological systems. In: Walter H, Brooks DE, Srere PA, editors, Microcompartmentation and Phase Separation in Cytoplasm, San Diego: Academic Press, volume 192 of *Int. Rev. Cytol.* pp. 281 – 302.

[15] Völker J, Breslauer KJ (2005) Communication between noncontacting macromolecules. Annu Rev Biophys Biomolec Struct 34:21 – 42.

[16] Päuser S, Zschunke A, Khuen A, Keller K (1995) Estimation of water content and water mobility in the nucleus and cytoplasm of *Xenopus laevis* oocytes by NMR spectroscopy. Magn Reson Imaging 13:269 – 276.

[17] Nakashima H, Nishikawa K (1994) Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. J Mol Biol 238:54 – 61.

[18] Cedano J, Aloy P, Pérez-Pons JA, Querol E (1997) Relation between amino acid composition and cellular location of proteins. J Mol Biol 266:594 – 600.

[19] Andrade MA, O'Donoghue SI, Rost B (1998) Adaptation of protein surfaces to subcellular location. J Mol Biol 276:517 – 525.

[20] Chou KC, Elrod DW (1999) Protein subcellular location prediction. Protein Eng 12:107 – 118.

[21] Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, et al. (2002) Subcellular localization of the yeast proteome. Genes Dev 16:707 – 719.

[22] Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, et al. (2003) Global analysis of protein localization in budding yeast. Nature 425:686 – 691.

[23] Pai HV, Starke DW, Lesnefsky EJ, Hoppel CL, Mieyal JJ (2007) What is the functional significance of the unique location of glutaredoxin 1 (GRx1) in the intermembrane space of mitochondria? Antioxid Redox Signal 9:2027 – 2033.

[24] Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao YM, et al. (2000) The yeast nuclear pore complex: Composition, architecture, and transport mechanism. J Cell Biol 148:635–651.

[25] Graham LA, Powell B, Stevens TH (2000) Composition and assembly of the yeast vacuolar H^+-ATPase complex. J Exp Biol 203:61–70.

[26] Prokisch H, Scharfe C, Camp II DG, Xiao WZ, David L, et al. (2004) Integrative analysis of the mitochondrial proteome in yeast. PLoS Biol 2:795 – 804.

[27] Kristensen AR, Schandorff S, Høyer-Hansen M, Nielsen MO, Jäättelä M, et al. (2008) Ordered organelle degradation during starvation-induced autophagy. Mol Cell Proteomics (in press).

[28] Häussinger D, Lang F, Gerok W (1994) Regulation of cell function by the cellular hydration state. Am J Physiol 267:E343 – E355.

[29] Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11:4241 – 4257.

[30] Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, et al. (2001) Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell 12:323 – 337.

[31] Chen DR, Toone WM, Mata J, Lyne R, Burns G, et al. (2003) Global transcriptional responses of fission yeast to environmental stress. Mol Biol Cell 14:214 – 229.

[32] Gibson BR, Lawrence SJ, Leclaire JPR, Powell CD, Smart KA (2007) Yeast responses to stresses associated with industrial brewery handling. Fems Microbiol Rev 31:535 – 569.

[33] Sella G, Hirsh AE (2005) The application of statistical physics to evolutionary biology. Proc Natl Acad Sci U S A 102:9541–9546.

[34] Aita T, Husimi Y (1996) Fitness spectrum among random mutants on Mt. Fuji-type fitness landscape. J Theor Biol 182:469–485.

[35] Jørgensen SE, Fath BD (2004) Application of thermodynamic principles in ecology. Ecological Complexity 1:267 – 280.

[36] Demetrius L, Ziehe M (2007) Darwinian fitness. Theor Popul Biol 72:323–345.

[37] Reiss J (2007) Relative fitness, teleology, and the adaptive landscape. Evol Biol 34:4–27.

[38] Battley EH (1992) Calculation of the thermodynamic properties of protein in *Escherichia coli* K-12 grown on succinic acid, energy changes accompanying protein anabolism, and energetic role of ATP in protein synthesis. Biotechnol Bioeng 40:280 – 288.

[39] Ferry JG, House CH (2006) The stepwise evolution of early life driven by energy conservation. Mol Biol Evol 23:1286 – 1292.

[40] Wächtershäuser G (1988) Before enzymes and templates: Theory of surface metabolism. Microbiol Rev 52:452 – 484.

[41] Shock EL, Schulte MD (1998) Organic synthesis during fluid mixing in hydrothermal systems. J Geophys Res - Planets 103:28513 – 28527.

[42] Mel HC, Ewald DA (1974) Thermodynamic potentials and evolution towards the stationary state in open systems of far-from-equilibrium chemical reactions: The affinity squared minimum function. J Math Biol 1:133 – 151.

[43] Schneider ED, Kay JJ (1994) Life as a manifestation of the second law of thermodynamics. Math Comput Model 19:25 – 48.

[44] Wicken JS (1979) Generation of complexity in evolution: A thermodynamic and information-theoretical discussion. J Theor Biol 77:349–365.

[45] Wright BE (2004) Stress-directed adaptive mutations and evolution. Mol Microbiol 52:643–650.

[46] Fong SS, Joyce AR, Palsson BO (2005) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. Genome Res 15:1365–1372.

[47] Wicken JS (1980) A thermodynamic theory of evolution. J Theor Biol 87:9–23.

[48] Lv J, Li N, Niu DK (2008) Association between the availability of environmental resources and the atomic composition of organismal proteomes: Evidence from *Prochlorococcus* strains living at different depths. Biochem Biophys Res Comm 375:241–246.

[49] Schekman R (1985) Protein localization and membrane traffic in yeast. Annu Rev Cell Biol 1:115–143.

[50] Gunning P, Weinberger R, Jeffrey P, Hardeman E (1998) Isoform sorting and the creation of intracellular compartments. Annu Rev Cell Dev Biol 14:339–372.

[51] Doxsey S, McCollum D, Theurkauf W (2005) Centrosomes in cellular regulation. Annu Rev Cell Dev Biol 21:411–434.

[52] Schoenheimer R, Ratner S, Rittenberg D (1939) Studies in protein metabolism X. The metabolic activity of body proteins investigated with *l* (–)-leucine containing two isotopes. J Biol Chem 130:703 – 732.

[53] Swick RW (1958) Measurement of protein turnover in rat liver. J Biol Chem 231:751 – 764.

[54] Halvorson H (1958) Intracellular protein and nucleic acid turnover in resting yeast cells. Biochim Biophys Acta 27:255 – 266.

[55] Berg JM, Tymoczko JL, Stryer L (2007) Biochemistry. New York: W. H. Freeman and Company, 5th edition, 974 pp.

[56] Borsook H, Dubnoff JW (1940) The biological synthesis of hippuric acid in vitro. J Biol Chem 132:307 – 324.

[57] Morowitz HJ (1978) Foundations of Bioenergetics. New York: Academic Press, 344 pp.

[58] Seligmann H (2003) Cost-minimization of amino acid usage. J Mol Evol 56:151 – 161.

[59] Swire J (2007) Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. J Mol Evol 64:558 – 571.

[60] Zeldovich KB, Berezovsky IN, Shakhnovich EI (2007) Protein and DNA sequence determinants of thermophilic adaptation. PLoS Comput Biol 3:62 – 72. arxiv:q-bio/0607004.

[61] Korzhinskii DS (1966) On thermodynamics of open systems and the phase rule (A reply to D. F. Weill and W. S. Fyfe). Geochim Cosmochim Acta 30:829 – 835.

[62] Eugster HP, Skippen GB (1967) Igneous and metamorphic reactions involving gas equilibria. In: Abelson PH, editor, Researches in Geochemistry, New York: Wiley, volume 2. pp. 492 – 520.

[63] Helgeson HC (1970) A chemical and thermodynamic model of ore deposition in hydrothermal systems. In: Morgan BA, editor, Mineralogical Society of America, Fiftieth Anniversary Symposium, Mineralogical Society of America, volume Special Paper #3. pp. 155 – 186.

[64] Haas JR, Shock EL, Sassani DC (1995) Rare earth elements in hydrothermal systems: Estimates of standard partial molal thermodynamic properties of aqueous complexes of the rare earth elements at high pressures and temperatures. Geochim Cosmochim Acta 59:4329 – 4350.

[65] Seewald JS (2001) Aqueous geochemistry of low molecular weight hydrocarbons at elevated temperatures and pressures: Constraints from mineral buffered laboratory experiments. Geochim Cosmochim Acta 65:1641–1664.

[66] Helgeson HC, Richard L, McKenzie WF, Norton DL, Schmitt AM (2008) A chemical and thermodynamic model of oil generation in hydrocarbon source rocks. Geochim Cosmochim Acta (in press).

[67] Helgeson HC (1974) Chemical interaction of feldspars and aqueous solutions. In: MacKenzie WS, Zussman J, editors, The Feldspars, Manchester, England: Manchester University Press, chapter 11. pp. 184 – 217.

[68] Helgeson HC (1979) Mass transfer among minerals and hydrothermal solutions. In: Barnes HL, editor, Geochemistry of Hydrothermal Ore Deposits, New York: Wiley. 2nd edition, pp. 568 – 610.

[69] Dick JM, LaRowe DE, Helgeson HC (2006) Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins. Biogeosciences 3:311 – 336.

[70] Dick JM (2008) Calculation of the relative metastabilities of proteins using the CHNOSZ software package. Geochem Trans 9:10.

[71] Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, et al. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31:365 – 370.

[72] Pedrajas JR, Porras P, Martínez-Galisteo E, Padilla CA, Miranda-Vizuete A, et al. (2002) Two isoforms of *Saccharomyces cerevisiae* glutaredoxin 2 are expressed *in vivo* and localize to different subcellular compartments. Biochem J 364:617 – 623.

[73] Molina MM, Bellí G, de la Torre MA, Rodríguez-Manzaneque MT, Herrero E (2004) Nuclear monothiol glutaredoxins of *Saccharomyces cerevisiae* can function as mitochondrial glutaredoxins. J Biol Chem 279:51923 – 51930.

[74] Herrero E, Ros J, Tamarit J, Belli G (2006) Glutaredoxins in fungi. Photosynth Res 89:127 – 140.

[75] Pedrajas JR, Kosmidou E, Miranda-Vizuete A, Gustafsson JA, Wright APH, et al. (1999) Identification and functional characterization of a novel mitochondrial thioredoxin system in *Saccharomyces cerevisiae*. J Biol Chem 274:6366 – 6373.

[76] Garrels RM (1960) Mineral Equilibria. New York: Harper & Brothers, 254 pp.

[77] Dahm LJ, Jones DP (2000) Rat jejunum controls luminal thiol-disulfide redox. J Nutr 130:2739 – 2745.

[78] Drakulic T, Temple MD, Guido R, Jarolim S, Breitenbach M, et al. (2005) Involvement of oxidative stress response genes in redox homeostasis, the level of reactive oxygen species, and ageing in *Saccharomyces cerevisiae*. FEMS Yeast Res 5:1215 – 1228.

[79] Trotter EW, Grant CM (2003) Non-reciprocal regulation of the redox state of the glutathione-glutaredoxin and thioredoxin systems. EMBO Rep 4:184 – 188.

[80] Macville M, Schröck E, Padilla-Nash H, Keck C, Ghadimi BM, et al. (1999) Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. Cancer Res 59:141 – 150.

[81] ATCC: The Global Bioresource Center (2008). Product Description; CRL-1606. URL `http://www.atcc.org/common/catalog/numSearch/numResults.cfm?atccNum=CRL-1606`.

[82] Shock EL, Sassani DC, Willis M, Sverjensky DA (1997) Inorganic species in geologic fluids: Correlations among standard molal thermodynamic properties of aqueous ions and hydroxide complexes. Geochim Cosmochim Acta 61:907 – 950.

[83] Mojaverian P (1996) Evaluation of gastrointestinal pH and gastric residence time via the Heidelberg Radiotelemetry Capsule: Pharmaceutical application. Drug Dev Res 38:73–85.

[84] Chan P, Lovrić J, Warwicker J (2006) Subcellular pH and predicted pH-dependent features of proteins. Proteomics 6:3494 – 3501.

[85] Singh A, Kaur N, Kosman DJ (2007) The metalloreductase Fre6p in Fe-Efflux from the yeast vacuole. J Biol Chem 282:28619–28626.

[86] Go YM, Jones DP (2008) Redox compartmentalization in eukaryotic cells. Biochim Biophys Acta-Gen Subj 1780:1271–1290.

[87] Botstein D, Amberg D, Mulholland J, Huffaker T, Adams A, et al. (1997) The yeast cytoskeleton. In: Pringle JR, Broach JR, Jones EW, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology, New York: Cold Spring Harbor Laboratory Press. pp. 1 – 90.

[88] Alberts B, Bray D, Lewis J, Raff M, Roberts K, et al. (1989) Molecular Biology of the Cell. New York: Garland Publishing, Inc., 2nd edition, 1219 pp.

[89] Palache C (1929) Paragenetic classification of the minerals of Franklin, New Jersey. Am Miner 14:1 – 18.

[90] Nývlt J (1995) The Ostwald rule of stages. Cryst Res Technol 30:443 – 449.

[91] Steinmann P, Lichtner PC, Shotyk W (1994) Reaction path approach to mineral weathering reactions. Clay Clay Min 42:197 – 206.

[92] Lew DJ, Weinert T, Pringle JR (1997) Cell cycle control in *Saccharomyces cerevisiae*. In: Pringle JR, Broach JR, Jones EW, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology, New York: Cold Spring Harbor Laboratory Press. pp. 607 – 695.

[93] Wente SR, Gasser SM, Caplan AJ (1997) The nucleus and nucleocytoplasmic transport in *Saccharomyces cerevisiae*. In: Pringle JR, Broach JR, Jones EW, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology, New York: Cold Spring Harbor Laboratory Press. pp. 471 – 546.

[94] Kaiser CA, Gimeno RE, Shaywitz DA (1997) Protein secretion, membrane biogenesis, and endocytosis. In: Pringle JR, Broach JR, Jones EW, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology, New York: Cold Spring Harbor Laboratory Press. pp. 91 – 227.

[95] Jones EW, Webb GC, Hiller MA (1997) Biogenesis and function of the yeast vacuole. In: Pringle JR, Broach JR, Jones EW, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology, New York: Cold Spring Harbor Laboratory Press. pp. 363 – 470.

[96] Lazarow PB, Kunau W (1997) Peroxisomes. In: Pringle JR, Broach JR, Jones EW, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology, New York: Cold Spring Harbor Laboratory Press. pp. 547 – 605.

[97] Pon L, Schatz G (1991) Biogenesis of yeast mitochondria. In: Broach JR, Pringle JR, Jones EW, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Genome Dynamics, Protein Synthesis, and Energetics, New York: Cold Spring Harbor Laboratory Press. pp. 333 – 406.

[98] Wilson DF, Erecinska M, Dutton PL (1974) Thermodynamic relationships in mitochondrial oxidative phosphorylation. Annu Rev Biophys Bioeng 3:203–230.

[99] Veech RL (1978) Regulation of coenzyme potential by near equilibrium reactions. In: Srere PA, Estabrook RW, editors, Microenvironments and Metabolic Compartmentation, New York: Academic Press. pp. 17–64.

[100] Schafer FQ, Buettner GR (2001) Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. Free Radic Biol Med 30:1191 – 1212.

[101] Lloyd D, Murray DB (2006) The temporal architecture of eukaryotic growth. FEBS Lett 580:2830–2835.

[102] Menon SG, Goswami PC (2007) A redox cycle within the cell cycle: ring in the old with the new. Oncogene 26:1101 – 1109.

[103] D'Angiolella V, Santarpia C, Grieco D (2007) Oxidative stress overrides the spindle checkpoint. Cell Cycle 6:576–579.

[104] Murray DB (2004) On the temporal self-organisation of *Saccharomyces cerevisae*. Curr Genomics 5:665–671.

[105] Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, et al. (2003) Global analysis of protein expression in yeast. Nature 425:737 – 741.

[106] Helgeson HC (1987) Application of moderation theorems to metasomatic processes. In: Helgeson HC, editor, Chemical Transport in Metasomatic Processes, Dordrecht, Holland: D. Reidel. pp. 189 – 238.

[107] Prigogine I, Defay R (1954) Chemical Thermodynamics. London: Longmans, Green and Co.

[108] Helgeson HC, Kirkham DH (1974) Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: I. Summary of the thermodynamic/electrostatic properties of the solvent. Am J Sci 274:1089 – 1198.

[109] Wagman DD, Evans WH, Parker VB, Schumm RH, Halow I, et al. (1982) The NBS tables of chemical thermodynamic properties. Selected values for inorganic and C_1 and C_2 organic substances in SI units. J Phys Chem Ref Data 11:1 – 392.

[110] Shock EL, Helgeson HC, Sverjensky DA (1989) Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species. Geochim Cosmochim Acta 53:2157 – 2183.

[111] SGD Project (2007). *Saccharomyces* Genome Database. `http://www.yeastgenome.org`.

[112] Welch MD, Iwamatsu A, Mitchison TJ (1997) Actin polymerization is induced by Arp2/3 protein complex at the surface of *Listeria monocytogenes*. Nature 385:265–269.

[113] Miranda JJL, De Wulf P, Sorger PK, Harrison SC (2005) The yeast DASH complex forms closed rings on microtubules. Nat Struct Mol Biol 12:138–143.

[114] Mullins RD, Pollard TD (1999) Structure and function of the Arp2/3 complex. Curr Opin Struct Biol 9:244–249.

[115] Schmid M, Jaedicke A, Du TG, Jansen RP (2006) Coordination of endoplasmic reticulum and mRNA localization to the yeast bud. Curr Biol 16:1538–1543.

[116] Vinh DBN, Kern JW, Hancock WO, Howard J, Davis TN (2002) Reconstitution and characterization of budding yeast γ-tubulin complex. Mol Biol Cell 13:1144–1157.

[117] Frazier JA, Wong ML, Longtine MS, Pringle JR, Mann M, et al. (1998) Polymerization of purified yeast septins: Evidence that organized filament arrays may not be required for septin function. J Cell Biol 143:737–749.

[118] Conibear E, Cleck JN, Stevens TH (2003) Vps51p mediates the association of the GARP (Vps52/53/54) complex with the late Golgi t-SNARE Tlg1p. Mol Biol Cell 14:1610–1623.

[119] Mo CQ, Bard M (2005) Erg28p is a key protein in the yeast sterol biosynthetic enzyme complex. J Lipid Res 46:1991–1998.

[120] Bernstein KA, Gallagher JEG, Mitchell BM, Granneman S, Baserga SJ (2004) The small-subunit processome is a ribosome assembly intermediate. Eukaryot Cell 3:1619–1626.

[121] Burri L, Lithgow T (2004) A complete set of SNAREs in yeast. Traffic 5:45–52.

[122] Kostelansky MS, Schluter C, Tam YYC, Lee S, Ghirlando R, et al. (2007) Molecular architecture and functional model of the complete yeast ESCRT-I heterotetramer. Cell 129:485–498.

[123] Hierro A, Sun J, Rusnak AS, Kim J, Prag G, et al. (2004) Structure of the ESCRT-II endosomal trafficking complex. Nature 431:221–225.

[124] Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440:631–636.

[125] Tuller T, Kupiec M, Ruppin E (2007) Determinants of protein abundance and translation efficiency in *S. cerevisiae*. PLoS Comput Biol 3:2510–2519.

[126] Wicken JS (1987) Evolution, Thermodynamics, and Information. Oxford University Press, 243 pp.

[127] Drever JI (1997) The Geochemistry of Natural Waters. Upper Saddle River, New Jersey: Prentice Hall, 3rd edition.

[128] Hendrickson JB, Cram DJ, Hammond GS (1970) Organic Chemistry. New York: McGraw-Hill, 3rd edition, 1279 pp.

[129] Buvet R (1983) General criteria for the fulfillment of redox reactions. In: Milazzo G, Blank M, editors, Bioelectrochemistry I: Biological Redox Reactions. New York: Plenum Press, volume 11 of *Ettore Majorana International Science Series*, pp. 15 – 50.